Central limit theorem and bootstrap procedure for Wasserstein's variations with an application to structural relationships between distributions

Eustasio del Barrio, Paula Gordaliza, Hélène Lescornel, Jean-Michel Loubes

Central limit theorem and bootstrap procedure for Wasserstein's variations with an application to structural relationships between distributions

Eustasio del Barrio[a], Paula Gordaliza[a], Hélène Lescornel[b] and Jean-Michel Loubes[b1]
[a]*IMUVA, Universidad de Valladolid* and [b]*Institut de mathématiques de Toulouse*

Abstract

Wasserstein barycenters and variance-like criteria based on the Wasserstein distance are used in many problems to analyze the homogeneity of collections of distributions and structural relationships between the observations. We propose the estimation of the quantiles of the empirical process of Wasserstein's variation using a bootstrap procedure. We then use these results for statistical inference on a distribution registration model for general deformation functions. The tests are based on the variance of the distributions with respect to their Wasserstein's barycenters for which we prove central limit theorems, including bootstrap versions.

Keywords: Central Limit Theorem, Goodness-of-fit, Wasserstein distance

1. Introduction

Analyzing the variability of large data sets is a difficult task when the inner geometry of the information conveyed by the observations is far from being Euclidean. Indeed, deformations on the data such as location-scale transformations or more general warping procedures preclude the use of common statistical methods. Looking for a way to measure structural relationships within data is of high importance. Such issues arise when considering the estimation of probability measures observed with deformations; it is common, e.g., when considering gene expression.

Over the last decade, there has been a large amount of work dealing with registrations issues. We refer, e.g., to [3, 5, 28] and references therein. However, when dealing with the registration of warped distributions, the literature is scarce. We mention here the method provided for biological computational issues known as quantile normalization in [9, 21] and references therein. Recently, using optimal transport methodologies, comparisons of distributions have been studied using a notion of Fréchet mean for distributions as in [1] or a notion of depth as in [10].

As a natural frame for applications of a deformation model, consider J independent random samples of size n, where for each $j \in \{1, \dots, J\}$, the real-valued random variable X_j has distribution μ_j and, for each $i \in \{1, \dots, n\}$, the ith observation of X_j is such that

$$X_{i,j} = g_j(\varepsilon_{i,j}),$$

where the $\varepsilon_{i,j}$s are iid random variables with unknown distribution μ. Assume that the functions $g_1, \dots, g_J$ belong to a class $\mathcal{G}$ of deformation functions, which model how the distributions $\mu_1, \dots, \mu_J$ are warped one to another.

This model is the natural extension of the functional deformation models studied in the statistical literature for which estimation procedures are provided in [22] and testing issues are tackled in [11]. Note that at the era of parallelized inference where a large amount of data is processed in the same way but at different locations or by different computers, this framework appears also natural since this parallelization may lead to small changes with respect to the law of the observations that should be eliminated.

In the framework of warped distributions, a central goal is the estimation of the warping functions, possibly as a first step towards registration or alignment of the (estimated) distributions. Of course, without some constraints on the class $\mathcal{G}$, the deformation model is meaningless. We can, for instance, obtain any distribution on $\mathbb{R}^d$ as a warped version of a fixed probability having a density if we take the optimal transportation map as the warping function; see [34]. One has to consider smaller classes of deformation functions to perform a reasonable registration.

[1]Corresponding author: loubes@math.univ-toulouse.fr

In cases where $\mathcal{G}$ is a parametric class, estimation of the warping functions is studied in [2]. However, estimation/registration procedures may lead to inconsistent conclusions if the chosen deformation class $\mathcal{G}$ is too small. It is, therefore, important to be able to assess the fit to the deformation model given by a particular choice of $\mathcal{G}$. This is the main goal of this paper. We note that within this framework, statistical inference on deformation models for distributions has been studied first in [20]. Here we provide a different approach which allows to deal with more general deformation classes.

The pioneering works [15, 25] study the existence of relationships between distributions F and G by using a discrepancy measure $\Delta(F, G)$ between them which is built using the Wasserstein distance. The authors consider the assumption $\mathcal{H}_0 : \Delta(F, G) > \Delta_0$ versus $\mathcal{H}_a : \Delta(F, G) \leq \Delta_0$ for a chosen threshold Δ_0. Thus when the null hypothesis is rejected, there is statistical evidence that the two distributions are similar with respect to the chosen criterion. In this same vein, we define a notion of variation of distributions using the Wasserstein distance, W_r, in the set $\mathcal{W}_r(\mathbb{R}^d)$ of probability measures with finite rth moments, where $r \geq 1$. This notion generalizes the concept of variance for random distributions over $\mathbb{R}^d$. This quantity can be defined as

$$V_r(\mu_1, \ldots, \mu_J) = \inf_{\eta \in \mathcal{W}_r(\mathbb{R}^d)} \left\{ \frac{1}{J} \sum_{j=1}^{J} W_r^r(\mu_j, \eta) \right\}^{1/r},$$

which measures the spread of the distributions. Then, to measure closeness to a deformation model, we take a look at the minimal variation among warped distributions, a quantity that we could consider as a minimal alignment cost. Under some mild conditions, a deformation model holds if and only if this minimal alignment cost is null and we can base our assessment of a deformation model on this quantity.

As in [15, 25], we provide results (a Central Limit Theorem and bootstrap versions) that enable to reject that the minimal alignment cost exceeds some threshold, and hence to conclude that it is below that threshold. Our results are given in a setup of general, nonparametric classes of warping functions. We also provide results in the somewhat more restrictive setup where one is interested in the more classical goodness-of-fit problem for the deformation model. Note that a general Central Limit Theorem is available for the Wasserstein distance in [18].

The paper is organized as follows. The main facts about Wasserstein variation are presented in Section 2, together with the key idea that fit to a deformation model can be recast in terms of the minimal Wasserstein variation among warped versions of the distributions. In Section 3, we prove some Lipschitz bounds for the law of empirical Wasserstein variations as well as of minimal alignment costs on $\mathbb{R}^d$. As a consequence of these results, the quantiles of the minimal warped variation criterion can be consistently estimated by some suitable bootstrap quantiles, which can be approximated by simulation, yielding consistent tests of fit to deformation models, provided that the empirical criterion has a regular limiting distribution.

Central Limit Theorems for empirical minimal Wasserstein variation are further explored for univariate distributions in Section 4, which covers nonparametric deformation models, and in Section 5, which presents a sharper analysis for the case of semiparametric deformation models. These sections describe consistent tests for deformation models in the corresponding setups. Section 6 reports some simulations assessing the quality of the bootstrap procedure. Finally, proofs are gathered in the Appendix.

2. Wasserstein variation and deformation models for distributions

Much recent work has been conducted to measure the spread or the inner structure of a collection of distributions. In this paper, we define a notion of variability which relies on the notion of Fréchet mean for the space of probabilities endowed with the Wasserstein metrics, of which we will recall the definition hereafter. First, for any integer $d \geq 1$, consider the set $\mathcal{W}_r(\mathbb{R}^d)$ of probabilities with finite rth moment. For μ and ν in $\mathcal{W}_r(\mathbb{R}^d)$, we denote by $\Pi(\mu, \nu)$ the set of all probability measures π over the product set $\mathbb{R}^d \times \mathbb{R}^d$ with first (respectively second) marginal μ (respectively ν). The L_r transportation cost between these two measures is defined as

$$W_r(\mu, \nu)^r = \inf_{\pi \in \Pi(\mu,\nu)} \int \|x - y\|^r \, d\pi(x, y).$$

This transportation cost makes it possible to endow the set $\mathcal{W}_r(\mathbb{R}^d)$ with the metric $W_r(\mu, \nu)$. More details on Wasserstein distances and their links with optimal transport problems can be found, e.g., in [26, 34].

Within this framework, we can define a global measure of separation of a collection of probability measures as follows. Given $\mu_1, \ldots, \mu_J \in \mathcal{W}_r(\mathbb{R}^d)$, let

$$V_r(\mu_1, \ldots, \mu_J) = \inf_{\eta \in \mathcal{W}_r(\mathbb{R}^d)} \left\{ \frac{1}{J} \sum_{j=1}^{J} W_r^r(\mu_j, \eta) \right\}^{1/r}$$

be the Wasserstein r-variation of $\mu_1, \ldots, \mu_J$ or the variance of the μ_js.

The special case $r = 2$ has been studied in the literature. The existence of a minimizer of the map $\eta \mapsto \{W_2^2(\mu_1, \eta) + \cdots + W_2^2(\mu_J, \eta)\}/J$ is proved in [1], as well as its uniqueness under some smoothness assumptions. Such a minimizer, μ_B, is called a barycenter or Fréchet mean of $\mu_1, \ldots, \mu_J$. Hence,

$$V_2(\mu_1, \ldots, \mu_J) = \left\{ \frac{1}{J} \sum_{j=1}^{J} W_2^2(\mu_j, \mu_B) \right\}^{1/2}.$$

Empirical versions of the barycenter are analyzed in [8, 24]. Similar ideas have also been developed in [6, 14].

This quantity, which is an extension of the variance for probability distributions is a good candidate to evaluate the concentration of a collection of measures around their Fréchet mean. In particular, it can be used to measure the fit to a distribution deformation model. More precisely, assume as in the Introduction that we observe J independent random samples with sample $j \in \{1, \ldots, J\}$ consisting of iid observations $X_{1,j}, \ldots, X_{n,j}$ with common distribution μ_j. We assume that $\mathcal{G}_j$ is a family (parametric or nonparametric) of invertible warping functions and denote $\mathcal{G} = \mathcal{G}_1 \times \cdots \times \mathcal{G}_J$. Then, the deformation model assumes that

$$\text{there exists } (\varphi_1^*, \ldots, \varphi_J^*) \in \mathcal{G} \text{ and iid } (\varepsilon_{i,j})_{1 \le i \le n, 1 \le j \le J} \text{ such that for all } j \in \{1, \ldots, J\}, X_{i,j} = (\varphi_j^*)^{-1}(\varepsilon_{i,j}). \tag{1}$$

Equivalently, the deformation model (1) means that there exists $(\varphi_1^*, \ldots, \varphi_J^*) \in \mathcal{G}$ such that the collection of $\varphi_j^*(X_{i,j})$s taken over all $j \in \{1, \ldots, J\}$ and $i \in \{1, \ldots, n\}$ is iid or, if we write $\mu_j(\varphi_j)$ for the distribution of $\varphi_j(X_{i,j})$, that there exists $(\varphi_1^*, \ldots, \varphi_J^*) \in \mathcal{G}$ such that $\mu_1(\varphi_1^*) = \cdots = \mu_J(\varphi_J^*)$.

We propose to use the Wasserstein variation to measure the fit of model (1) through the minimal alignment cost

$$A_r(\mathcal{G}) = \inf_{(\varphi_1, \ldots, \varphi_J) \in \mathcal{G}} V_r^r \{\mu_1(\varphi_1), \ldots, \mu_J(\varphi_J)\}. \tag{2}$$

Let us assume that $\mu_1(\varphi_1), \ldots, \mu_J(\varphi_J)$, $(\varphi_1, \ldots, \varphi_J) \in \mathcal{G}$ are in $\mathcal{W}_r(\mathbb{R}^d)$. If the deformation model (1) holds, then $A_r(\mathcal{G}) = 0$. Under the additional mild assumption that the minimum in (2) is attained, we have that the deformation model can be equivalently formulated as $A_r(\mathcal{G}) = 0$ and a goodness-of-fit test to the deformation model becomes, formally, a test of

$$\mathcal{H}_0 : A_r(\mathcal{G}) = 0 \quad \text{vs.} \quad \mathcal{H}_a : A_r(\mathcal{G}) > 0. \tag{3}$$

A testing procedure can be based on the empirical version of $A_r(\mathcal{G})$, namely,

$$A_{n,r}(\mathcal{G}) = \inf_{(\varphi_1, \ldots, \varphi_J) \in \mathcal{G}} V_r^r \{\mu_{n,1}(\varphi_1), \ldots, \mu_{n,J}(\varphi_J)\}, \tag{4}$$

where $\mu_{n,j}(\varphi_j)$ denotes the empirical measure on $\varphi_j(X_{1,j}), \ldots, \varphi_j(X_{n,j})$. We would reject the deformation model (1) for large values of $A_{n,r}(\mathcal{G})$.

As noted in [15, 25], the testing problem (3) can be considered as a mere sanity check for the deformation model, since lack of rejection of the null does not provide statistical evidence that the deformation model holds. Consequently, as in the cited references, we will also consider the alternative testing problem

$$\mathcal{H}_0 : A_r(\mathcal{G}) \ge \Delta_0 \quad \text{vs.} \quad \mathcal{H}_a : A_r(\mathcal{G}) < \Delta_0, \tag{5}$$

where $\Delta_0 > 0$ is a fixed threshold. With this formulation the test decision of rejecting the null hypothesis implies that there is statistical evidence that the deformation model is approximately true. In this case, rejection would correspond to small observed values of $A_{n,r}(\mathcal{G})$. In subsequent sections, we provide theoretical results that allow the computation of approximate critical values and p-values for the testing problems (3) and (5) under suitable assumptions.

3. Bootstraping Wasserstein's variations

We present now some general results on Wasserstein distances that will be applied to estimate the asymptotic distribution of the minimal alignment cost statistic, $A_{n,r}(\mathcal{G})$, defined in (4). In this section, we write $\mathcal{L}(Z)$ for the law of any random variable Z. We note the abuse of notation in the following, in which W_r is used both for the Wasserstein distance on $\mathbb{R}$ and on $\mathbb{R}^d$, but this should not cause much confusion.

Our first result shows that the laws of empirical transportation costs are continuous (and even Lipschitz) functions of the underlying distributions.

Theorem 1. *Set ν, ν', η probability measures in $\mathcal{W}_r(\mathbb{R}^d)$, $Y_1, \ldots, Y_n$ iid random vectors with common law ν, $Y'_1, \ldots, Y'_n$, iid with law ν' and write ν_n, ν'_n for the corresponding empirical measures. Then*

$$W_r[\mathcal{L}\{W_r(\nu_n, \eta)\}, \mathcal{L}\{W_r(\nu'_n, \eta)\}] \leq W_r(\nu, \nu').$$

The deformation assessment criterion introduced in Section 2 is based on the Wasserstein r-variation of distributions, V_r. It is convenient to note that $V_r^r(\nu_1, \ldots, \nu_J)$ can also be expressed as

$$V_r^r(\nu_1, \ldots, \nu_J) = \inf_{\pi \in \Pi(\nu_1, \ldots, \nu_J)} \int T(y_1, \ldots, y_J) d\pi(y_1, \ldots, y_J), \tag{6}$$

where $\Pi(\nu_1, \ldots, \nu_J)$ denotes the set of probability measures on $\mathbb{R}^d$ with marginals $\nu_1, \ldots, \nu_J$ and

$$T(y_1, \ldots, y_J) = \min_{z \in \mathbb{R}^d} \frac{1}{J} \sum_{j=1}^{J} \|y_j - z\|^r.$$

Here we are interested in empirical Wasserstein r-variations, namely, the r-variations computed from the empirical measures $\nu_{n_j,j}$ coming from independent samples $Y_{1,j}, \ldots, Y_{n_j,j}$ of iid random variables with distribution ν_j. Note that in this case, problem (6) is a linear optimization problem for which a minimizer always exists.

As before, we consider the continuity of the law of empirical Wasserstein r-variations with respect to the underlying probabilities. This is covered in the next result.

Theorem 2. *With the above notation,*

$$W_r^r[\mathcal{L}\{V_r(\nu_{n_1,1}, \ldots, \nu_{n_J,J})\}, \mathcal{L}\{V_r(\nu'_{n_1,1}, \ldots, \nu'_{n_J,J})\}] \leq \frac{1}{J} \sum_{j=1}^{J} W_r^r(\nu_j, \nu'_j).$$

A useful consequence of the above results is that empirical Wasserstein distances or r-variations can be bootstrapped under rather general conditions. To be more precise, in Theorem 1 we take $\nu' = \nu_n$, the empirical measure on $Y_1, \ldots, Y_n$, and consider a bootstrap sample $Y_1^*, \ldots, Y_{m_n}^*$ of iid (conditionally given $Y_1, \ldots, Y_n$) observations with common law ν_n. We will assume that the resampling size m_n satisfies $m_n \to \infty$, $m_n = o(n)$ and write $\nu^*_{m_n}$ for the empirical measure on $Y_1^*, \ldots, Y_{m_n}^*$ and $\mathcal{L}^*(Z)$ for the conditional law of Z given $Y_1, \ldots, Y_n$. Theorem 1 now reads

$$W_r[\mathcal{L}^*\{W_r(\nu^*_{m_n}, \nu)\}, \mathcal{L}(W_r\{\nu_{m_n}, \nu)\}] \leq W_r(\nu_n, \nu).$$

Hence, if $W_r(\nu_n, \nu) = O_{\Pr}(1/r_n)$ for some sequence $r_n > 0$ such that $r_{m_n}/r_n \to 0$ as $n \to \infty$, then, using the fact that $W_r\{\mathcal{L}(aX), \mathcal{L}(aY)\} = aW_r\{\mathcal{L}(X), \mathcal{L}(Y)\}$ for $a > 0$, we see that

$$W_r[\mathcal{L}^*\{r_{m_n} W_r(\nu^*_{m_n}, \nu)\}, \mathcal{L}\{r_{m_n} W_r(\nu_{m_n}, \nu)\}] \leq \frac{r_{m_n}}{r_n} r_n W_r(\nu_n, \nu) \to 0$$

in probability.

Assume that, in addition, $r_n W_r(\nu_n, \nu) \rightsquigarrow \gamma(\nu)$ for a smooth distribution $\gamma(\nu)$. If $\hat{c}_n(\alpha)$ denotes the αth quantile of the conditional distribution $\mathcal{L}^*\{r_{m_n} W_r(\nu^*_{m_n}, \nu)\}$, then

$$\lim_{n \to \infty} \Pr\{r_n W_r(\nu_n, \nu) \leq \hat{c}_n(\alpha)\} = \alpha; \tag{7}$$

see, e.g., Lemma 1 in [23]. We conclude in this case that the quantiles of $r_n W_r(\nu_n, \nu)$ can be consistently estimated by the bootstrap quantiles, $\hat{c}_n(\alpha)$, which, in turn, can be approximated through Monte Carlo simulation.

As an example, if $d = 1$ and $r = 2$, under integrability and smoothness assumptions on ν, we have

$$\sqrt{n}\, W_2(\nu_n, \nu) \rightsquigarrow \left[\int_0^1 \frac{B^2(t)}{f^2\{F^{-1}(t)\}} dt\right]^{1/2},$$

as $n \to \infty$, where f and F^{-1} are the density and the quantile function of ν, respectively; see [17]). Therefore, Eq. (7) holds. Bootstrap results have also been provided in [19].

For the deformation model (1), statistical inference is based on $A_{n,r}(\mathcal{G})$, introduced in (4). Now consider $A'_{n,r}(\mathcal{G})$, the corresponding version obtained from samples with underlying distributions μ'_j. Then, a version of Theorem 2 is valid for these minimal alignment costs, provided that the deformation classes are uniformly Lipschitz, namely, under the assumption that, for all $j \in \{1, \ldots, J\}$,

$$L_j = \sup_{x \neq y, \varphi_j \in \mathcal{G}_j} \frac{\|\varphi_j(x) - \varphi_j(x)\|}{\|x - y\|} \tag{8}$$

is finite.

Theorem 3. *If $L = \max(L_1, \ldots, L_j) < \infty$, with L_j as in* (8)*, then*

$$W_r^r\left[\mathcal{L}[\{A_{n,r}(\mathcal{G})\}^{1/r}], \mathcal{L}[\{A'_{n,r}(\mathcal{G})\}^{1/r}]\right] \leq L^r \frac{1}{J} \sum_{j=1}^{J} W_r^r(\mu_j, \mu'_j).$$

Hence, the Wasserstein distance of the variance of two collections of distributions can be controlled using the distance between the distributions. The main consequence of this fact is that the minimal alignment cost can also be bootstrapped as soon as a distributional limit theorem exists for $A_{n,r}(\mathcal{G})$, as in the discussion above. In Sections 4 and 5 below, we present distributional results of this type in the one-dimensional case. We note that, while general Central Limit Theorems for the empirical transportation cost are not available in dimension $d > 1$, some recent progress has been made in this direction; see, e.g., [29] for Gaussian distributions and [31], which gives such results for distributions on $\mathbb{R}^d$ with finite support. Further advances along these lines would make it possible to extend the results in the following section to higher dimensions.

4. Assessing the fit of nonparametric deformation models

In this section and subsequent ones, we focus on the case $d = 1$ and $r = 2$. Thus we will simply write $A(\mathcal{G})$ and $A_n(\mathcal{G})$ (instead of $A_2(\mathcal{G})$ and $A_{2,n}(\mathcal{G})$) for the minimal alignment cost and its empirical version, defined in (2) and (4). Otherwise we keep the notation from Section 2, with $X_{1,j}, \ldots, X_{n,j}$ iid random variables with law μ_j being one of the J independent samples. Now $\mathcal{G}_j$ is a class of invertible warping functions from $\mathbb{R}$ to $\mathbb{R}$, which we assume to be increasing. We note that in this case the barycenter of a set of probability measures $\mu_1, \ldots, \mu_J$ with distribution functions $F_1, \ldots, F_J$ is the probability having quantile function $F_B^{-1} = (F_1^{-1} + \cdots + F_J^{-1})/J$; see, e.g., [1]. We observe further that $\mu_j(\varphi_j)$ is determined by the quantile function $\varphi_j \circ F_j^{-1}$. We will write

$$F_B^{-1}(\varphi) = \frac{1}{J} \sum_{j=1}^{J} \varphi_j \circ F_j^{-1}$$

for the quantile function of the barycenter of $\mu_1(\varphi_1), \ldots, \mu_J(\varphi_J)$, while $\rightsquigarrow$ will denote convergence in distribution.

In order to prove a Central Limit Theorem for $A_n(\mathcal{G})$, we need to make assumptions on the integrability and regularity of the distributions $\mu_1, \ldots, \mu_J$ as well as on the smoothness of the warping functions. We consider first the assumptions on the distributions. For each $j \in \{1, \ldots, J\}$, the distribution function associated with μ_j is denoted

F_j. We will assume that μ_j is supported on a (possibly unbounded) interval in the interior of which F_j is C^2 and $F_j' = f_j > 0$ and satisfies

$$\sup_x \frac{F_j(x)\{1 - F_j(x)\}|f_j'(x)|}{f_j(x)^2} < \infty, \tag{9}$$

and, further, that for some $q > 1$,

$$\int_0^1 \frac{\{t\,(1-t)\}^{q/2}}{[f_j\{F_j^{-1}(t)\}]^q} dt < \infty \tag{10}$$

and for some $r > 4$,

$$\mathrm{E}(|X_j|^r) < \infty. \tag{11}$$

Assumption (9) is a classical regularity requirement for the use of strong approximations for the quantile process, as in [13, 17]. Our proof relies on the use of these techniques. As for (10) and (11), they are mild integrability conditions. If F_j has regularly varying tails of order $-r$ (e.g., Pareto tails) then both conditions hold — and also (9) — so long as $r > 4$ and $1 < q < 2r/(r+2)$. Of course the conditions are fulfilled by distributions with lighter tails such as exponential or Gaussian laws for any $q \in (1, 2)$.

Turning to the assumptions on the classes of warping functions, we recall that a uniform Lipschitz condition was needed for the approximation bound in Theorem 3. For the Central Limit Theorem in this section, we need some refinement of that condition, the extent of which will depend on the integrability exponent q in (10), as follows. We set $p_0 = \max\{q/(q-1), 2\}$ and define on $\mathcal{H}_j = C^1(\mathbb{R}) \cap L^{p_0}(X_j)$ the norm

$$\|h_j\|_{\mathcal{H}_j} = \sup|h_j'(x)| + \mathrm{E}\{|h_j(X_j)|^{p_0}\}^{1/p_0},$$

and on the product space $\mathcal{H}_1 \times \cdots \times \mathcal{H}_J$, $\|h\|_{\mathcal{H}} = \|h_1\|_{\mathcal{H}_1} + \cdots + \|h_J\|_{\mathcal{H}_J}$. We further assume that

$$\mathcal{G}_j \subset \mathcal{H}_j \text{ is compact for } \|\cdot\|_{\mathcal{H}_j} \text{ and } \sup_{h\in\mathcal{G}_j} |h'(x_n^h) - h'(x)| \underset{\sup_{h\in\mathcal{G}_j}|x_n^h - x|\to 0}{\longrightarrow} 0, \tag{12}$$

and, finally, that for some $r > \max(4, p_0)$,

$$\mathrm{E}\sup_{h\in\mathcal{G}_j} |h(X_j)|^r < \infty. \tag{13}$$

We note that (13) is a slight strengthening of the uniform moment bound already contained in (12); we could take $p_0 > \max\{q/(q-1), 4\}$ in (12) and (13) would follow.

Our next result gives a Central Limit Theorem for $A_n(\mathcal{G})$ under the assumptions on the distributions and deformation classes described above. The limit can be simply described in terms of a centered Gaussian process indexed by the set of minimizers of the variation functional, namely, $U(\varphi) = V_2^2\{\mu_1(\varphi_1), \ldots, \mu_J(\varphi_J)\}$. An elementary computation shows that

$$\{U^{1/2}(\varphi) - U^{1/2}(\tilde{\varphi})\}^2 \le \frac{1}{J}\sum_{j=1}^J \mathrm{E}\{\varphi_j(X_j) - \tilde{\varphi}_j(X_j)\}^2,$$

from which we conclude continuity of U with respect to $\|\cdot\|_{\mathcal{H}}$. In particular, the set

$$\Gamma = \left\{\varphi \in \mathcal{G} :\ U(\varphi) = \inf_{\phi\in\mathcal{G}} U(\phi)\right\}$$

is a nonempty compact subset of $\mathcal{G}$.

Theorem 4. *Assume that $B_1, \ldots, B_J$ are mutually independent Brownian bridges. Set*

$$c_j(\varphi) = 2\int_0^1 \varphi_j' \circ F_j^{-1}\{\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)\}\frac{B_j}{f_j \circ F_j^{-1}}$$

and $C(\varphi) = \{c_1(\varphi) + \cdots + c_J(\varphi)\}/J$, $\varphi \in \mathcal{G}$. Then, under assumptions (9)–(13), C is a centered Gaussian process on $\mathcal{G}$ with trajectories that are almost surely continuous with respect to $\|\cdot\|_{\mathcal{H}}$. Furthermore,

$$\sqrt{n}\,\{A_n(\mathcal{G}) - A(\mathcal{G})\} \rightsquigarrow \min_{\varphi\in\Gamma} C(\varphi).$$

A proof of Theorem 4 is given in the Appendix. The random variables

$$\int_0^1 \varphi_j' \circ F_j^{-1} \frac{B_j}{f_j \circ F_j^{-1}} \{\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)\}$$

are centered Gaussian, with variance

$$\int_{[0,1]^2} \{\min(s,t) - st\} \frac{\varphi_j'\{F_j^{-1}(t)\}}{f_j\{F_j^{-1}(t)\}} \left[\varphi_j\{F_j^{-1}(t)\} - F_B^{-1}(\varphi)(t)\right] \frac{\varphi_j'\{F_j^{-1}(s)\}}{f_j\{F_j^{-1}(s)\}} \left[\varphi_j\{F_j^{-1}(s)\} - F_B^{-1}(\varphi)(s)\right] ds dt.$$

In particular, if U has a unique minimizer the limiting distribution in Theorem 4 is Gaussian. However, our result works in more generality, even without uniqueness assumptions.

We remark also that although we have focused for simplicity on the case of samples of equal size, the case of different sample sizes $n_1, \ldots, n_J$ can also be handled with straightforward changes. More precisely, let us write $A_{n_1,\ldots,n_J}(\mathcal{G})$ for the minimal alignment cost computed from the empirical distribution of the samples and assume that $n_j \to \infty$ and

$$\frac{n_j}{n_1 + \cdots + n_J} \to (\gamma_j)^2 > 0,$$

then with straightforward changes in our proof we can see that

$$\sqrt{\frac{n_1 \ldots n_J}{(n_1 + \cdots + n_J)^{J-1}}} \{A_{n_1,\ldots,n_J}(\mathcal{G}) - A(\mathcal{G})\} \rightsquigarrow \min_{\varphi \in \Gamma} \tilde{C}(\varphi),$$

where $\tilde{C}(\varphi) = \{\tilde{c}_1(\varphi) + \cdots + \tilde{c}_J(\varphi)\}/J$ and $\tilde{c}_j(\varphi) = (\Pi_{p \neq j} \gamma_p) c_j(\varphi)$.

If we try, as argued in Section 2, to base on $A_n(\mathcal{G})$ our assessment of fit to the deformation model (1), we should note that the limiting distribution in Theorem 4 depends on the unknown distributions μ_j and cannot be used for the computation of approximate critical values or p-values without further adjustments. We show now how this can be done in the case of the testing problem (5), namely, the test of

$$\mathcal{H}_0 : A_r(\mathcal{G}) \geq \Delta_0 \quad \text{vs.} \quad \mathcal{H}_a : A_r(\mathcal{G}) < \Delta_0,$$

for some fixed threshold $\Delta_0 > 0$, through the use of a bootstrap procedure.

Let us consider bootstrap samples $X_{1,j}^*, \ldots, X_{m_n,j}^*$ of iid observations sampled from $\mu_{n,j}$, the empirical distribution on $X_{1,j}, \ldots, X_{n,j}$. We write $\mu_{m_n,j}^*$ for the empirical measure on $X_{1,j}^*, \ldots, X_{m_n,j}^*$ and introduce

$$A_{m_n}^*(\mathcal{G}) = \inf_{\varphi \in \mathcal{G}} V_2^2\{\mu_{m_n,1}^*(\varphi_1), \ldots, \mu_{m_n,J}^*(\varphi_J)\}.$$

Now, we base our testing procedure on the conditional α-quantiles (given the $X_{i,j}$s) of $\sqrt{m_n}\{A_{m_n}^*(\mathcal{G}) - \Delta_0\}$, which we denote $\hat{c}_n(\alpha; \Delta_0)$. Our next result, which follows from Theorems 3 and 4, shows that the test that rejects $\mathcal{H}_0$ when

$$\sqrt{n}\{A_n(\mathcal{G}) - \Delta_0\} < \hat{c}_n(\alpha; \Delta_0)$$

is a consistent test of approximate level α for (5). We note that the bootstrap quantiles $\hat{c}_n(\alpha; \Delta_0)$ can be computed using Monte Carlo simulation.

Corollary 5. *If $m_n \to \infty$, and $m_n = O(\sqrt{n})$, then under assumptions (9)–(13)*

$$\Pr\left[\sqrt{n}\{A_n(\mathcal{G}) - \Delta_0\} < \hat{c}_n(\alpha; \Delta_0)\right] \to \begin{cases} 0 & \text{if } A(\mathcal{G}) > \Delta_0, \\ \alpha & \text{if } A(\mathcal{G}) = \Delta_0, \\ 1 & \text{if } A(\mathcal{G}) < \Delta_0. \end{cases}$$

Rejection in the testing problem (5) would result, as noted in Section 2, in statistical evidence supporting that the deformation model holds approximately, and hence that related registration methods can be safely applied. If, nevertheless, we were interested in gathering statistical evidence against the deformation model, then we should consider the classical goodness-of-fit problem (3). Some technical difficulties arise then. Note that if the deformation model holds, that is, if $A(\mathcal{G}) = 0$, then we have $\varphi_j \circ F_j^{-1} = F_B^{-1}(\varphi)$ for each $\varphi \in \Gamma$, which implies that the result of Theorem 4 becomes $\sqrt{n}\, A_n(\mathcal{G}) \rightsquigarrow 0$. Hence, a nondegenerate limit law for $A_n(\mathcal{G})$ in this case requires a more refined analysis, that we handle in the next section.

5. Goodness-of-fit in semiparametric deformation models

In many cases, deformation functions can be made more specific in the sense that they follow a known shape depending on parameters that may differ for sample to sample. In our approach to the classical goodness-of-fit problem (3), we consider a parametric model in which $\varphi_j = \varphi_{\theta_j}$ for some finite-dimensional parameter θ_j that describes the warping effect within a fixed shape. Now, that the deformation model holds means that there exist $\theta^* = (\theta_1^*, \ldots, \theta_J^*)$ such that for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, J\}$, $X_{i,j} = \varphi_{\theta_j^*}^{-1}(\varepsilon_{i,j})$. Hence, from now on, we will consider the following family of deformations, indexed by a parameter $\lambda \in \Lambda \subset \mathbb{R}^p$:

$$\varphi : \Lambda \times \mathbb{R} \to \mathbb{R} : (\lambda, x) \mapsto \varphi_\lambda(x).$$

The classes $\mathcal{G}_j$ become now $\{\varphi_{\theta_j} : \theta_j \in \Lambda\}$. We denote $\Theta = \Lambda^J$ and write $A_n(\Theta)$ and $A(\Theta)$ instead of $A_n(\mathcal{G})$ and $A(\mathcal{G})$. We also use the simplified notation $\mu_j(\theta_j)$ instead of $\mu_j(\varphi_{\theta_j})$, $F_B(\theta)$ for $F_B(\varphi_{\theta_1}, \ldots, \varphi_{\theta_J})$ and similarly for the empirical versions. Our main goal is to identify a weak limit theorem for $A_n(\Theta)$ under the null in (3). Therefore, throughout this section, we assume that model (1) holds. This means, in particular, that the quantile functions of the samples satisfy $F_j^{-1} = \varphi_{\theta_j^*}^{-1} \circ G^{-1}$, with G the distribution function of the $\varepsilon_{i,j}$s. As before, we assume that the warping functions are invertible and increasing, which now means that, for each $\lambda \in \Lambda$, φ_λ is an invertible, increasing function.

It is convenient at this point to introduce the notation

$$\psi_j(\lambda, x) = \varphi_\lambda\{\varphi_{\theta_j^*}^{-1}(x)\}$$

for all $j \in \{1, \ldots, J\}$ and ε for a random variable with the same distribution as the $\varepsilon_{i,j}$. Note that $\psi_j(\theta_j^*, x) = x$.

Now, under smoothness assumptions on the functions ψ_j that we present in detail below, if the parameter space is compact then the function

$$U_n(\theta_1, \ldots, \theta_J) = V_2^2\{\mu_{n,1}(\theta_1), \ldots, \mu_{n,J}(\theta_J)\}$$

admits a minimizer that we will denote by $\hat{\theta}_n$, i.e.,

$$\hat{\theta}_n \in \operatorname*{argmin}_{\theta \in \Theta} U_n(\theta).$$

Of course, since we are assuming that the deformation model holds, we know that θ^* is a minimizer of

$$U(\theta_1, \ldots, \theta_J) = V_2^2\{\mu_1(\theta_1), \ldots, \mu_J(\theta_J)\}.$$

For a closer analysis of the asymptotic behavior of $A_n(\Theta)$ under the deformation model, we need to make the following identifiability assumption

$$\theta^* \text{ belongs to the interior of } \Lambda \text{ and is the unique minimizer of } U. \tag{14}$$

Note that, equivalently, this means that θ^* is the unique zero of U.

As in the case of nonparametric deformation models, we need to impose some conditions on the class of warping functions and on the distribution of the errors, the $\varepsilon_{i,j}$s. For the former, we write D or D_u for derivative operators with respect to parameters. Hence, for instance, $D\psi_j(\lambda, x) = (D_1\psi_j(\lambda, x), \ldots, D_p\psi_j(\lambda, x))^\top$ is the vector consisting of partial derivatives of ψ_j with respect to its first p arguments evaluated at (λ, x); $D^2\psi_j(\lambda, x) = (D_{u,v}\psi_j(\lambda, x))_{u,v}$ is the hessian matrix for fixed x and so on. In what follows, $\psi_j'(\lambda, x)$ and similar notation will stand for derivatives with respect to x. Then we will assume that for each $j \in \{1, \ldots, J\}$, $u, v \in \{1, \ldots, p\}$, and some $r > 4$

$$\psi_j(\cdot, \cdot) \text{ is } C^2, \tag{15}$$

$$\mathrm{E}\Big\{\sup_{\lambda \in \Lambda} |\psi_j(\lambda, \varepsilon)|^r\Big\} < \infty, \quad \mathrm{E}\Big\{\sup_{\lambda \in \Lambda} |D_u\psi_j(\lambda, \varepsilon)|^r\Big\} < \infty, \quad \mathrm{E}\Big\{\sup_{\lambda \in \Lambda} |D_{u,v}\psi_j(\lambda, \varepsilon)|^r\Big\} < \infty, \tag{16}$$

and

$$\psi_j'(\cdot, \cdot) \text{ is bounded on } \Lambda \times \mathbb{R} \text{ and } \sup_{\lambda \in \Lambda} \left|\psi_j'(\lambda, x_n^\lambda) - \psi_j'(\lambda, x)\right| \xrightarrow{\sup_{\lambda \in \Lambda} |x_n^\lambda - x| \to 0} 0. \tag{17}$$

Turning to the distribution of the errors, we will assume that G is C^2 with $G'(x) = g(x) > 0$ on some interval and

$$\sup_x \frac{G(x)\{1 - G(x)\}|g'(x)|}{g(x)^2} < \infty. \tag{18}$$

Additionally (but see the comments after Theorem 6 below), we make the assumption that

$$\int_0^1 \frac{t(1-t)}{g^2\{G^{-1}(t)\}} dt < \infty. \tag{19}$$

Finally, before stating the asymptotic result for $A_n(\Theta)$, we introduce the $p \times p$ matrices

$$\Sigma_{i,i} = \frac{2(J-1)}{J^2} \int_0^1 D_i\psi_i\{\theta_i^*, G^{-1}(t)\}\psi_i\{\theta_i^*, G^{-1}(t)\}^\top dt,$$

$$\Sigma_{i,j} = -\frac{2}{J^2} \int_0^1 D_i\psi_i\{\theta_i^*, G^{-1}(t)\}\psi_i\{\theta_j^*, G^{-1}(t)\}^\top dt, \quad i \neq j$$

and the $pJ \times pJ$ matrix

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \cdots & \Sigma_{1,J} \\ \vdots & & \vdots \\ \Sigma_{J,1} & \cdots & \Sigma_{J,J} \end{bmatrix}.$$

This matrix is symmetric and positive semidefinite. To see this, consider $x_1, \ldots, x_J \in \mathbb{R}^p$ and $x^\top = [x_1^\top, \ldots, x_J^\top]$. Note that

$$\begin{aligned} x'\Sigma x &= \frac{2}{J^2} \int_0^1 \Big[\sum_i (J-1) \Big[x_i \times D_i\psi_i\{\theta_i^*, G^{-1}(t)\} \Big]^2 - 2 \sum_{i<j} \Big[x_i \times D_i\psi_i\{\theta_i^*, G^{-1}(t)\} \Big] \Big[x_j \times D_j\psi_j\{\theta_j^*, G^{-1}(t)\} \Big] \Big] dt \\ &= \frac{2}{J^2} \int_0^1 \sum_{i<j} \Big[[x_i \times D_i\psi_i\{\theta_i^*, G^{-1}(t)\}] - [x_j \times D_j\psi_j\{\theta_j^*, G^{-1}(t)\}] \Big]^2 dt \geq 0. \end{aligned}$$

In fact, Σ is positive definite, hence invertible, apart from some degenerate cases, For instance, if $p = 1$, Σ is invertible unless all the functions $D_i\psi_i\{\theta_i^*, G^{-1}(t)\}$ are proportional.

We are ready now for the announced distributional limit theorem.

Theorem 6. *Assume that the deformation model holds. Under assumptions* (14)–(18) $\hat{\theta}_n \to \theta^*$ *in probability. If, in addition,* Φ *is invertible, then* $\sqrt{n}(\hat{\theta}_n - \theta^*) \rightsquigarrow \Sigma^{-1}Y$, *where* $Y = (Y_1^\top, \ldots, Y_J^\top)^\top$ *with*

$$Y_j = \frac{2}{J} \int_0^1 D\psi_j\{\theta_j^*, G^{-1}(t)\} \frac{\tilde{B}_j(t)}{g\{G^{-1}(t)\}} dt,$$

$\tilde{B}_j = B_j - (B_1 + \cdots + B_J)/J$ *and* $B_1, \ldots, B_J$ *mutually independent Brownian bridges. Furthermore, if* (19) *also holds, then*

$$nA_n(\Theta) \rightsquigarrow \frac{1}{J} \sum_{j=1}^J \int_0^1 \Big(\frac{\tilde{B}_j}{g \circ G^{-1}} \Big)^2 - \frac{1}{2} Y^\top \Sigma^{-1} Y.$$

We make a number of comments here. First, we note that, while, for simplicity, we have formulated Theorem 6 assuming that the deformation model holds, the Central Limit Theorem for $\hat{\theta}_n$ still holds (with some additional assumptions and changes in Φ) in the case when the model is false and θ^* is not the true parameter, but the one that gives the best (but imperfect) alignment. Given that our focus here is the assessment of the deformation models, we refrain from pursuing this issue.

Our second comment is about the identifiability condition (14). At first sight it can seem to be too strong to be realistic. Actually, for some deformation models it could happen that $\varphi_\theta \circ \varphi_\eta = \varphi_{\theta * \eta}$ for some $\theta * \eta \in \Theta$. In this case,

if $X_{i,j} = \varphi^{-1}_{\theta^*_j}(\varepsilon_{i,j})$ with $\varepsilon_{i,j}$ iid, then, for any θ, $X_{i,j} = \varphi^{-1}_{\theta * \theta^*_j}(\tilde{\varepsilon}_{i,j})$ with $\tilde{\varepsilon}_{i,j} = \varphi_\theta(\varepsilon_{i,j})$ which are also iid; consequently, $(\theta * \theta^*_1, \ldots, \theta * \theta^*_J)$ is also a zero of U. This applies, for instance, to location and scale models.

A simple fix to this issue is to select one of the signals as the reference, say the Jth signal, and assume that θ^*_J is known since it can be, in fact, chosen arbitrarily. The criterion function becomes then $\tilde{U}(\theta_1, \ldots, \theta_{J-1}) = U(\theta_1, \ldots, \theta_{J-1}, \theta^*_J)$. One could then make the (more realistic) assumption that $\tilde{\theta}^* = (\theta^*_1, \ldots, \theta^*_{J-1})$ is the unique zero of $\tilde{U}$ and base the analysis on $\tilde{U}_n(\theta_1, \ldots, \theta_{J-1}) = U_n(\theta_1, \ldots, \theta_{J-1}, \theta^*_J)$ and $\hat{\tilde{\theta}}_n = \arg\min_{\tilde{\theta}} \tilde{U}_n(\tilde{\theta})$. The results in this section can be adapted almost verbatim to this setup. In particular,

$$\sqrt{n}\,(\hat{\tilde{\theta}}_n - \tilde{\theta}^*) \rightsquigarrow \tilde{\Sigma}^{-1}\tilde{Y},$$

with $\tilde{Y}^\top = (Y_1^\top, \ldots, Y_{J-1}^\top)$ and $\tilde{\Sigma} = (\Sigma_{i,j})_{1 \le i,j \le J-1}$. Again, the invertibility of $\tilde{\Sigma}$ is almost granted. In fact, arguing as above, we see that and $\tilde{\Sigma}$ is positive definite if the function $D\psi_i\{\theta^*_i, G^{-1}(t)\}$ is not null for all $i \in \{1, \ldots, J-1\}$.

Next, we discuss about the smoothness and integrability conditions on the errors. As before, (18) is a regularity condition that enables to use strong approximations for the quantile process. One might be surprised that the moment condition (11) does not show up here, but in fact it is contained in (16); recall that $\psi_j(\theta^*_j, x) = x$. The integrability condition (19) is necessary and sufficient for ensuring

$$\int_0^1 \frac{B(t)^2}{g^2\{G^{-1}(t)\}} dt < \infty,$$

from which we see that the limiting random variable in the last claim in Theorem 6 is an almost surely finite random variable. This implies that, as $n \to \infty$,

$$nW_2^2(G_n, G) \rightsquigarrow \int_0^1 \frac{B(t)^2}{g^2\{G^{-1}(t)\}} dt,$$

with G_n the empirical distribution function on a sample of size n and distribution function G. We refer to [17, 30] for details. Condition (11) is a strong assumption on the tails of G and does not include, e.g., Gaussian distributions. In contrast, under the less stringent condition

$$\int_0^1 \int_0^1 \frac{(s \wedge t - st)^2}{g^2\{G^{-1}(s)\} g^2\{G^{-1}(t)\}} ds dt < \infty, \tag{20}$$

which is satisfied for Gaussian laws, it can be shown that the limit as $\delta \to 0$

$$\int_\delta^{1-\delta} \frac{B(t)^2 - t(1-t)}{g^2\{G^{-1}(t)\}} dt,$$

exists in probability and can be expressed as a weighted sum of independent, centered χ^2_1 random variables; see [17] for details. Then, denoting that kind of limit as

$$\int_0^1 \frac{B(t)^2 - t(1-t)}{g^2\{G^{-1}(t)\}} dt,$$

under some additional tail conditions — still satisfied by Gaussian distributions; these are conditions (2.10) and (2.22) to (2.24) in the cited reference — we have that, as $n \to \infty$,

$$nW_2^2(G_n, G) - c_n \rightsquigarrow \int_0^1 \frac{B(t)^2 - t(1-t)}{g^2\{G^{-1}(t)\}} dt,$$

with

$$c_n = \int_{1/n}^{1-1/n} \frac{EB(t)^2}{g^2\{G^{-1}(t)\}} dt.$$

A simple look at the proof of Theorem 5.1 shows that under these conditions, instead of (19), we can conclude that, as $n \to \infty$,

$$nA_n(\Theta) - (J-1)c_n/J^2 \rightsquigarrow \frac{1}{J}\sum_{j=1}^{J}\int_0^1 \frac{\tilde{B}_j^2(t) - (J-1)t(1-t)/J}{g^2\{G^{-1}(t)\}}dt - \frac{1}{2}Y^\top \Sigma^{-1} Y. \tag{21}$$

Our last comment about the assumptions for Theorem 6 concerns the compactness assumption on the parameter space. This may lead in some examples to artificial constraints on the parameter space. However, under some conditions (see, e.g., Corollary 3.2.3 in [33]) it is possible to prove that the global minimizer of the empirical criterion lies in a compact neighborhood of the true minimizer. In such cases the conclusion of Theorem 6 would extend to the unconstrained deformation model.

As a toy example consider the case of deformations by changes in scale, with $J = 2$. As above we fix the parameters of, say, the first sample, and consider the family of deformations $\varphi_\sigma(x) = \sigma x$. We assume that the deformation model holds, with the first sample having distribution function G and the second $(\sigma^*)^{-1}G^{-1}$; hence, σ^* is the unique minimizer of $U(\sigma)$. We find that

$$U_n(\sigma) = \int_0^1 (F_{n,1}^{-1} - \sigma F_{n,2}^{-1})^2/4,$$

from which we see that almost surely,

$$\hat{\sigma}_n = \left(\int F_{n,1}^{-1}F_{n,2}^{-1}\right) \Big/ \left\{\int (F_{n,2}^{-1})^2\right\} \to \sigma^*$$

and thus the conclusion of Theorem 6 remains valid if we take $\Theta = (0, \infty)$. To avoid further technicalities, we prefer to think of this as a different problem that should be handled in an ad hoc way for each particular example.

Turning back to our goal of assessment of the deformation model (1) based on the observed value of $A_n(\Theta)$, Theorem 6 gives some insight into the threshold levels for rejection of the null in the testing problem (3). However, the limiting distribution still depends on unknown objects and designing a tractable test requires to estimate the quantiles of this distribution. This is the goal of our next result.

We consider bootstrap samples $X_{1,j}^*, \ldots, X_{m_n,j}^*$ of iid observations sampled from μ_j^n, write $\mu_{m_n,j}^*$ for the empirical measure on $X_{1,j}^*, \ldots, X_{m_n,j}^*$ and $A_{m_n}^*(\Theta)$ for the minimal alignment cost computed from the bootstrap samples. We also write $\hat{c}_n(\alpha)$ for the conditional α quantile of $m_n A_{m_n}^*(\Theta)$ given the $X_{i,j}$.

Corollary 7. *Assume that the semiparametric deformation models holds. If $m_n \to \infty$, and $m_n/n \to 0$, then under assumptions* (14)–(19)*, we have that*

$$\Pr\{nA_n(\Theta) > \hat{c}_n(1-\alpha)\} \to \alpha. \tag{22}$$

Corollary 7 show that the test that rejects $\mathcal{H}_0 : A(\Theta) = 0$ (which, as disussed in section 2, is true if and only if the deformation model holds) when $nA_n(\Theta) > \hat{c}_n(1-\alpha)$ is asymptotically of level α. It is easy to check that the test is consistent against alternatives that satisfy regularity and integrability assumptions as in Theorem 6.

The key to Corollary 7 is that under the assumptions a bootstrap Central Limit Theorem holds for $m_n A_{m_n}^*(\Theta)$. As with Theorem 6, the integrability conditions on the errors can be relaxed and still have a bootstrap Central Limit Theorem. That would be the case if we replace (22) by (20) and the additional conditions mentioned above under which (21) holds. Then, the further assumption that the errors have a log-concave distribution and $m_n = O(n^\rho)$ for some $\rho \in (0, 1)$ would be enough to prove a bootstrap Central Limit Theorem, see the comments after the proof of Corollary 7 in the Appendix. In particular, a bootstrap Central Limit Theorem holds for Gaussian tails.

6. Simulations

We present in this section different simulations in order to study the goodness of fit test we propose in this paper. In this framework, we consider the scale-location family of deformations, i.e., $\theta^* = (\mu^*, \sigma^*)$ and observations such that $X_{i,j} = \mu_j^* + \sigma_j^* \epsilon_{i,j}$ for different distributions of $\epsilon_{i,j}$.

6.1. Construction of an α-level test

First, we aim at studying the bootstrap procedure which enables to build the test. For this we choose a level $\alpha = 0.05$ and aim at estimating the quantile of the asymptotic distribution using a bootstrap method.

Let B be the number of bootstrap samples, we proceed as follows to design a bootstrapped goodness of fit test.

1. For all $b \in \{1, \dots, B\}$,
   1.1. For $j \in \{1, \dots, J\}$, create a bootstrap sample $X_{1,j}^{*^b}, \dots, X_{m,j}^{*^b}$, with fixed size $m \in \{1, \dots, n\}$ of the first observation sample $X_{1,j}, \dots, X_{n,j}$,
   1.2. Compute $(u_m^{*b})^2 = \inf_{\theta \in \Theta} U_m^{*b}(\theta)$.
2. Sort the values $(u_m^{*b})^2$ for $b \in \{1, \dots, B\}$, viz. $(u_m^{*(1)})^2 \leqslant \cdots \leqslant (u_m^{*(B)})^2$, then take $\hat{q}_m(1-\alpha) = u_m^{*(B(1-\alpha))}$, the quantile of order $1-\alpha$ of the bootstrap distribution of the statistic $\inf_{\theta \in \Theta} U_n(\theta)$.
3. The test rejects the null hypothesis if $nu_n^2 > m[u_m^{*\{B(1-\alpha)\}}]^2$.

Once the test is built, we first ensure that the level of the test has been correctly achieved. For this we repeat the test for large K (here $K = 1000$) to estimate the probability of rejection of the test as

$$\hat{p}_n = \frac{1}{K} \sum_{k=1}^{K} \mathbf{1}\left[nu_{n,k}^2 > m \left[u_{m,k}^{*\{B(1-\alpha)\}} \right]^2 \right],$$

where $\mathbf{1}$ denotes an indicator function. We present in Table 1 these results for different J and several choices for $m = m_n$ depending on the size of the initial sample.

As expected, the bootstrap method makes it possible to build a test of level α provided the bootstrap sample is large enough. The required size of the sample increases with the number of different distributions J to be tested.

6.2. Power of the test procedure

Then we compute the power of previous test for several situations. In particular we must compute the probability of rejection of the null hypothesis under $\mathcal{H}_a$. Hence for several distributions, we test the assumption that the model comes from a warping frame, when observations from a different distribution called γ are observed. The simulations are conducted for the following choices of sample size and for the different distributions:

$J = 2$: $\mathcal{N}(0, 1)$ and γ;

$J = 3$:$\mathcal{N}(0, 1)$, $\mathcal{N}(5, 2^2)$ and γ;

$J = 5$: $\mathcal{N}(0, 1)$, $\mathcal{N}(5, 2^2)$, $\mathcal{N}(3, 1)$, $\mathcal{N}(1.5, 3^2)$ and γ;

$J = 10$: $\mathcal{N}t(0, 1)$, $\mathcal{N}(5, 2^2\ )$, $\mathcal{N}(3, 1\ )$, $\mathcal{N}(1.5, 3^2)$, $\mathcal{N}(7, 4^2)$, $\mathcal{N}(2.5, 0.5^2)$, $\mathcal{N}(1, 1.5^2)$, $\mathcal{N}(4, 3^2)$, $\mathcal{N}(6, 5^2)$ and γ;

and also for different choices of γ.

Exponential distribution with parameter 1, $\mathcal{E}(1)$;

Double exponential (or Laplace) with parameter 1;

Student's t distribution $t_{(3)}$ and $t_{(4)}$ with 3 and 4 degrees of freedom.

All simulations were done for different sample sizes and different bootstrap samples, n and m_n. The results are presented in Tables 2, 3, 4 and 5, respectively.

We observe that the power of the test is very high in most cases. For the Exponential distribution, the power is close to 1. Indeed this distribution is very different from the Gaussian distribution since it is asymmetric, making it easy to discard the null assumption. The three other distributions do share with the Gaussian the property of symmetry, and yet the power of the test is also close to 1; it also increases with the sample size. Finally, for Student's t distribution, the higher the number of degrees of freedom, the more similar it becomes to a Gaussian distribution. This explains why it becomes more difficult for the test to reject the null hypothesis when using a Student $t_{(4)}$ rather than a $t_{(3)}$.

Appendix

A. Proofs of the results in Section 3

Proof of Theorem 1. We set $T_n = W_r(\nu_n, \eta)$, $T'_n = W_r(\nu'_n, \eta)$, and let $\Pi_n(\eta)$ be the set of probabilities on $\{1, \ldots, n\} \times \mathbb{R}^d$ with first marginal equal to the discrete uniform distribution on $\{1, \ldots, n\}$ and second marginal equal to η. We note that $T_n = \inf_{\pi \in \Pi_n(\eta)} a(\pi)$ if we denote

$$a(\pi) = \left\{ \int_{\{1,\ldots,n\} \times \mathbb{R}^d} \|Y_i - z\|^r d\pi(i, z) \right\}^{1/r}.$$

We define similarly $a'(\pi)$ from the Y'_i sample to get $T'_n = \inf_{\pi \in \Pi_n(\eta)} a'(\pi)$. But then, using the inequality $|\,\|a\| - \|b\|\,| \leq \|a - b\|$, we get

$$|a(\pi) - a'(\pi)| \leq \left\{ \int_{\{1,\ldots,n\} \times \mathbb{R}^d} \|Y_i - Y'_i\|^r d\pi(i, z) \right\}^{1/r} = \left(\frac{1}{n} \sum_{i=1}^n \|Y_i - Y'_i\|^r \right)^{1/r}.$$

This implies that

$$|T_n - T'_n|^r \leq \frac{1}{n} \sum_{i=1}^n \|Y_i - Y'_i\|^r.$$

If we take now (Y, Y') to be an optimal coupling of ν and ν', so that $\mathrm{E}(\|Y - Y'\|^r) = W_r^r(\nu, \nu')$ and $(Y_1, Y'_1), \ldots, (Y_n, Y'_n)$ to be iid copies of (Y, Y'), we see that for the corresponding realizations of T_n and T'_n, we have

$$\mathrm{E}(|T_n - T'_n|^r) \leq \frac{1}{n} \sum_{i=1}^n \mathrm{E}\left(\|Y_i - Y'_i\|^r\right) = W_r(\nu, \nu')^r.$$

But this shows that $W_r\{\mathcal{L}(T_n), \mathcal{L}(T'_n)\} \leq W_r(\nu, \nu')$, as claimed. □

Proof of Theorem 2. We write $V_{r,n} = V_r(\nu_{n_1,1}, \ldots, \nu_{n_J,J})$ and $V'_{r,n} = V_r(\nu'_{n_1,1}, \ldots, \nu'_{n_J,J})$. We note that

$$V_{r,n}^r = \inf_{\pi \in \Pi(U_1,\ldots,U_J)} \int T(i_1, \ldots, i_J) d\pi(i_1, \ldots, i_J),$$

where U_j is the discrete uniform distribution on $\{1, \ldots, n_j\}$ and

$$T(i_1, \ldots, i_J) = \min_{z \in \mathbb{R}^d} \frac{1}{J} \sum_{j=1}^J \|Y_{i_j,j} - z\|^r.$$

We write $T'(i_1, \ldots, i_J)$ for the equivalent function computed from the $Y'_{i,j}$s. Hence we have

$$|T(i_1, \ldots, i_J)^{1/r} - T'(i_1, \ldots, i_J)^{1/r}|^r \leq \frac{1}{J} \sum_{j=1}^J \|Y_{i_j,j} - Y'_{i_j,j}\|^r,$$

which implies

$$\begin{aligned}
&\left| \left\{ \int T(i_1, \ldots, i_J) d\pi(i_1, \ldots, i_J) \right\}^{1/r} - \left\{ \int T(i_1, \ldots, i_J) d\pi(i_1, \ldots, i_J) \right\}^{1/r} \right|^r \\
&\quad \leq \int \frac{1}{J} \sum_{j=1}^J \|Y_{i_j,j} - Y'_{i_j,j}\|^r d\pi(i_1, \ldots, i_J) \\
&\quad = \frac{1}{J} \sum_{j=1}^J \int \|Y_{i_j,j} - Y'_{i_j,j}\|^r d\pi(i_1, \ldots, i_J) = \frac{1}{J} \sum_{j=1}^J \left(\frac{1}{n_j} \sum_{i=1}^{n_j} \|Y_{i,j} - Y'_{i,j}\|^r \right).
\end{aligned}$$

Therefore,

$$|V_{r,n} - V'_{r,n}|^r \leq \frac{1}{J}\sum_{j=1}^{J}\left(\frac{1}{n_j}\sum_{i=1}^{n_j}\|Y_{i,j} - Y'_{i,j}\|^r\right).$$

If we take (Y_j, Y'_j) to be an optimal coupling of ν_j and ν'_j and $(Y_{1,j}, Y'_{1,j}), \ldots, (Y_{n_j,j}, Y'_{n_j,j})$ to be iid copies of (Y_j, Y'_j) for all $j \in \{1, \ldots, J\}$, then we obtain

$$\mathrm{E}\left(|V_{r,n} - V'_{r,n}|^r\right) \leq \frac{1}{J}\sum_{j=1}^{J}\left\{\frac{1}{n_j}\sum_{i=1}^{n_j}\mathrm{E}\left(\|Y_{i,j} - Y'_{i,j}\|^r\right)\right\} = \frac{1}{J}\sum_{j=1}^{J} W_r^r(\nu_j, \nu'_j).$$

The conclusion follows. □

Proof of Theorem 3. We argue as in the proof of Theorem 2 and write

$$A_{n,r}(\mathcal{G}) = \inf_{\varphi\in\mathcal{G}}\left\{\inf_{\pi\in\Pi(U_1,\ldots,U_J)}\int T(\varphi; i_1, \ldots, i_J)d\pi(i_1, \ldots, i_J)\right\},$$

where

$$T(\varphi; i_1, \ldots, i_J) = \min_{y\in\mathbb{R}}\frac{1}{J}\sum_{j=1}^{J}\|Z_{i_j,j}(\varphi_j) - y\|^r.$$

We write $T'(\varphi; i_1, \ldots, i_J)$ for the same function computed on the $Z'_{i,j}(\varphi_j)$'s. Now, from the fact $\|Z_{i,j}(\varphi_j) - Z'_{i,j}(\varphi_j)\|^r \leq L^r\|X_{i,j} - X'_{i,j}\|^r$ we see that

$$|T(\varphi; i_1, \ldots, i_J)^{1/r} - T'(\varphi; i_1, \ldots, i_J)^{1/r}|^r \leq L^r\frac{1}{J}\sum_{j=1}^{J}\|X_{i_j,j} - X'_{i_j,j}\|^r$$

and, as a consequence,

$$|V_r\{\mu_{n,1}(\varphi_1), \ldots, \mu_{n,J}(\varphi_J)\} - V_r\{\mu'_{n,1}(\varphi_1), \ldots, \mu'_{n,J}(\varphi_J)\}|^r \leq \frac{L^r}{J}\sum_{j=1}^{J}\sum_{i_j=1}^{n_j}\frac{1}{n_j}\|X_{i_j,j} - X'_{i_j,j}\|^r$$

which implies

$$|\{A_{n,r}(\mathcal{G})\}^{1/r} - \{A'_{n,r}(\mathcal{G})\}^{1/r}|^r \leq \frac{L^r}{J}\sum_{j=1}^{J}\left(\frac{1}{n_j}\sum_{i=1}^{n_j}\|X_{i,j} - X'_{i,j}\|^r\right).$$

If, as in the proof of Theorem 2, we assume that $(X_{i,j}, X'_{i,j})$ with $i \in \{1, \ldots, n_j\}$ are iid copies of an optimal coupling for μ_j and μ'_j, with different samples independent from each other we obtain that

$$\mathrm{E}\left[|\{A_{n,r}(\mathcal{G})\}^{1/r} - \{A'_{n,r}(\mathcal{G})\}^{1/r}|^r\right] \leq \frac{L^r}{J}\sum_{j=1}^{J} W_r^r(\mu_j, \mu'_j).$$

This concludes the argument. □

6.3. Proofs for results from Sections 4 and 5

We provide here proofs of the main results in Sections 4 and 5. Our approach relies on the consideration the processes defined, for all $\varphi \in \mathcal{G}$, by

$$C_n(\varphi) = \sqrt{n}\{U_n(\varphi) - U(\varphi)\} \quad \text{and} \quad C(\varphi) = \frac{1}{J}\sum_{j=1}^{J} c_j(\varphi),$$

where $U_n(\varphi) = V_2^2\{\mu_{n,1}(\varphi_1), \ldots, \mu_{n,J}(\varphi_J)\}$, $U(\varphi) = V_2^2\{\mu_1(\varphi_1), \ldots, \mu_J(\varphi_J)\}$,

$$c_j(\varphi) = 2\int_0^1 \varphi_j' \circ F_j^{-1}\{\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)\}\frac{B_j}{f_j \circ F_j^{-1}}$$

and $B_1, \ldots, B_J$ are independent standard Brownian bridges on $(0, 1)$. We prove below that the empirical deformation cost process C_n converges weakly to C as random elements in $L^\infty(\mathcal{G})$, the space of bounded, real-valued functions on $\mathcal{G}$. Theorem 4 will follow as a corollary of this result.

We will make frequent use in this section of the following technical lemma, which follows easily from the triangle and Hölder's inequalities. We omit the proof.

Lemma 8. *Under Assumption* (13)*,*

(i)

$$\sup_{\varphi_j \in \mathcal{G}_j} \sqrt{n}\int_0^{1/n}\left(\varphi_j \circ F_j^{-1}\right)^2 \to 0, \quad \sup_{\varphi_j \in \mathcal{G}_j} \sqrt{n}\int_{1-1/n}^1\left(\varphi_j \circ F_j^{-1}\right)^2 \to 0.$$

(ii)

$$\sup_{\varphi_j \in \mathcal{G}_j} \sqrt{n}\int_0^{1/n}\left(\varphi_j \circ F_{n,j}^{-1}\right)^2 \to 0, \quad \sup_{\varphi_j \in \mathcal{G}_j} \sqrt{n}\int_{1-1/n}^1\left(\varphi_j \circ F_{n,j}^{-1}\right)^2 \to 0$$

in probability.

(iii) If moreover (10) *holds, then for all* $j, k \in \{1, \ldots, J\}$*,*

$$\int_0^1 \frac{\sqrt{t(1-t)}}{f_k\{F_k^{-1}(t)\}} \sup_{\varphi_j \in \mathcal{G}_j}\left|\varphi_j\{F_j^{-1}(t)\}\right| dt < \infty.$$

Theorem 9. *Under assumptions* (9)–(13)*, C_n and C have almost surely trajectories in $L^\infty(\mathcal{G})$. Furthermore, C is a tight Gaussian random elemnt and C_n converges weakly to C in $L^\infty(\mathcal{G})$.*

Proof. We start by noting that

$$U_n(\varphi) = \frac{1}{J}\sum_{j=1}^J \int_0^1 \left\{\varphi_j \circ F_{n,j}^{-1} - F_{n,B}^{-1}(\varphi)\right\}^2, \quad U(\varphi) = \frac{1}{J}\sum_{j=1}^J \int_0^1 \left\{\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)\right\}^2$$

with

$$F_{n,B}^{-1}(\varphi) = \frac{1}{J}\sum_{j=1}^J \varphi_j \circ F_{n,j}^{-1}, \quad F_B^{-1}(\varphi) = \frac{1}{J}\sum_{j=1}^J \varphi_j \circ F_j^{-1}.$$

Now, (13) implies that

$$\sup_{\varphi_j \in \mathcal{G}_j} \int_0^1 \left(\varphi_j \circ F_j^{-1}\right)^2 < \infty.$$

Similarly, assumption (12) implies $K_j = \sup_{\varphi_j \in \mathcal{G}_j, x \in (c_j, d_j)} |\varphi_j'(x)| < \infty$. Noting that

$$\int_0^1 \left(\varphi_j \circ F_{n,j}^{-1}\right)^2 \le 2\int_0^1 \left(\varphi_j \circ F_j^{-1}\right)^2 + 2K_j^2 \int_0^1 \left(F_{n,j}^{-1} - F_j^{-1}\right)^2,$$

we see that

$$\sup_{\varphi_j \in \mathcal{G}_j} \int_0^1 \left(\varphi_j \circ F_{n,j}^{-1}\right)^2 < \infty, \text{ a.s.}$$

and, with little additional effort, conclude that C_n has almost surely bounded trajectories. Furthermore, writing

$$d_{j,k}(\varphi) = \int_0^1 \varphi_j' \circ F_j^{-1} \frac{B_j}{f_j \circ F_j^{-1}} \varphi_k \circ F_k^{-1},$$

we see that for $\varphi, \rho \in \mathcal{G}$

$$\begin{aligned}
&|d_{j,k}(\varphi) - d_{j,k}(\rho)| \\
&\quad \le \|\varphi_j' - \rho_j'\|_\infty \left| \int_0^1 \frac{B_k}{f_k \circ F_k^{-1}} \varphi_k \circ F_k^{-1} \right| + \left| \int_0^1 \rho_j' \circ F_j^{-1} \frac{B_k}{f_k \circ F_k^{-1}} \left(\varphi_k \circ F_k^{-1} - \rho_k \circ F_k^{-1} \right) \right| \\
&\quad \le \|\varphi_j' - \rho_j'\|_\infty \sup_{\varphi_k \in \mathcal{G}_k} \left| \int_0^1 \frac{B_k}{f_k \circ F_k^{-1}} \varphi_k \circ F_k^{-1} \right| + \sup_{(c_j, d_j)} |\rho_j'| \left(\int_0^1 \left| \frac{B_k}{f_k \circ F_k^{-1}} \right|^q \right)^{1/q} \left(\int_0^1 \left| \varphi_k \circ F_k^{-1} - \rho_k \circ F_k^{-1} \right|^{p_0} \right)^{1/p_0}
\end{aligned}$$

But using (iii) of Lemma 8, we deduce that

$$\mathrm{E}\left(\sup_{\varphi_k \in \mathcal{G}_k} \left| \int_0^1 \frac{B_k}{f_k \circ F_k^{-1}} \varphi_k \circ F_k^{-1} \right| \right) \le \int_0^1 \frac{\sqrt{t(1-t)}}{f_k\{F_k^{-1}(t)\}} \sup_{\varphi_j \in \mathcal{G}_j} |\varphi_j\{F_j^{-1}(t)\}| dt < \infty.$$

Hence, almost surely,

$$\sup_{\varphi \in \mathcal{G}} \left| \int_0^1 \frac{B_j}{f_j \circ F_j^{-1}} \varphi_j \circ F_j^{-1} \right| < \infty.$$

Furthermore, from assumption (10), we get that, almost surely,

$$\int_0^1 \left(\frac{B_j}{f_j \circ F_j^{-1}} \right)^q < \infty,$$

and thus, for some almost surely finite random variable T, $|d_{j,k}(\varphi) - d_{j,k}(\rho)| \le T\|\varphi - \rho\|_{\mathcal{G}}$ for $\varphi, \rho \in \mathcal{G}$. From this conclude that the trajectories of C are a.s. bounded, uniformly continuous functions on $\mathcal{G}$, endowed with the norm $\|\cdot\|_{\mathcal{G}}$ introduced in (12). In particular, C is a tight random element in $L^\infty(\mathcal{G})$; see, e.g., pp. 39–41 in [33].

From this point on, we pay attention to the quantile processes defined, for all $j \in \{1, \ldots, J\}$ and $t \in (0, 1)$, by

$$\rho_{n,j}(t) = \sqrt{n} f_j \left\{ F_j^{-1}(t) \right\} \left\{ F_{n,j}^{-1}(t) - F_j^{-1}(t) \right\}.$$

A trivial adaptation of Theorem 2.1 on p. 381 of [13] shows that, under (9), there exist, on a rich enough probability space, independent versions of $\rho_{n,j}$ and independent families of Brownian bridges $\{B_{n,j}\}_{n=1}^\infty$ for all $j \in \{1, \ldots, J\}$, satisfying

$$n^{1/2-\nu} \sup_{1/n \le t \le 1-1/n} \frac{|\rho_{n,j}(t) - B_{n,j}(t)|}{\{t(1-t)\}^\nu} = \begin{cases} O_p(\ln n) & \text{if } \nu = 0, \\ O_p(1) & \text{if } 0 < \nu \le 1/2. \end{cases} \tag{A.1}$$

We work, without loss of generality, with these versions of $\rho_{n,j}$ and $B_{n,j}$. We show now that

$$\sup_{\varphi \in \mathcal{G}} |C_n(\varphi) - \hat{C}_n(\varphi)| \to 0 \text{ in probability} \tag{A.2}$$

with

$$\hat{C}_n(\varphi) = \frac{1}{J} \sum_{j=1}^J c_{n,j}(\varphi) \quad \text{and} \quad c_{n,j}(\varphi) = 2 \int_0^1 \varphi_j' \circ F_j^{-1} \left\{ \varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi) \right\} \frac{B_{n,j}}{f_j \circ F_j^{-1}}.$$

To check this, we note that some simple algebra yields

$$C_n(\varphi) = \frac{2}{J} \sum_{j=1}^J \tilde{c}_{n,j} + \frac{1}{J} \sum_{j=1}^J \tilde{r}_{n,j}$$

with

$$\tilde{c}_{n,j} = \sqrt{n}\int_0^1 \left(\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1}\right)\left(\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)\right),$$

$$\tilde{r}_{n,j} = \sqrt{n}\int_0^1 \left[\left(\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1}\right) - \left\{F_{n,B}^{-1}(\varphi) - F_B^{-1}(\varphi)\right\}\right]^2.$$

From the elementary inequality $(a_1 + \cdots + a_J)^2 \le Ja_1^2 + \cdots + Ja_J^2$, we get that

$$\frac{1}{J}\sum_{j=1}^J \tilde{r}_{n,j} \le \frac{4\sqrt{n}}{J}\sum_{j=1}^J \int_0^1 \left(\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1}\right)^2 \le \frac{4\sqrt{n}}{J}\sum_{j=1}^J K_j \int_0^1 \left(F_{n,j}^{-1} - F_j^{-1}\right)^2,$$

with $K_j = \sup_{\varphi_j\in\mathcal{G}_j, x\in(c_j,d_j)} |\varphi_j'(x)| < \infty$, as above. Now we can use (11) and argue as in the proof of Theorem 2 in [4] to conclude that $\sqrt{n}\int_0^1 (F_{n,j}^{-1} - F_j^{-1})^2 \to 0$ in probability and, as a consequence, that

$$\sup_{\varphi\in\mathcal{G}} \left| C_n(\varphi) - \frac{1}{J}\sum_{j=1}^J \tilde{c}_{n,j}(\varphi) \right| \to 0 \text{ in probability.} \tag{A.3}$$

Furthermore, the Cauchy–Schwarz inequality implies that

$$n\left[\int_0^{1/n} \left(\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1}\right)\left\{\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)\right\}\right]^2$$
$$\le \sqrt{n}\int_0^{1/n} \left(\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1}\right)^2 \sqrt{n}\int_0^{1/n} \left\{\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)\right\}^2$$

and using (i) and (ii) of Lemma 8, the two factors converge to zero uniformly in φ. A similar argument works for the upper tail and allows to conclude that we can replace $\tilde{c}_{n,j}(\varphi)$ in (A.3) with

$$\tilde{\tilde{c}}_{n,j}(\varphi) = 2\sqrt{n}\int_{1/n}^{1-1/n} \left(\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1}\right)\left\{\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)\right\}.$$

Moreover,

$$\sup_{\varphi\in\mathcal{G}} \left| \int_0^{1/n} \varphi_j' \circ F_j^{-1} \frac{B_{n,j}}{f_j \circ F_j^{-1}} \left\{\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)\right\} \right| \le K_j \int_0^{1/n} \left| \frac{B_{n,j}}{f_j \circ F_j^{-1}} \right| \sup_{\varphi\in\mathcal{G}} \left| \left\{\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)\right\} \right|$$

and by (iii) of Lemma 8 and Cauchy–Schwarz's inequality,

$$\mathrm{E}\left[\int_0^{\frac{1}{n}} \left| \frac{B_{n,j}}{f_j \circ F_j^{-1}} \right| \sup_{\varphi\in\mathcal{G}} \left| \left\{\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)\right\} \right| \right] \le \int_0^{\frac{1}{n}} \frac{\sqrt{t(1-t)}}{f_j\{F_j^{-1}(t)\}} \sup_{\varphi\in\mathcal{G}} \left| \varphi_j\{F_j^{-1}(t)\} - F_B^{-1}(\varphi)(t) \right| dt \to 0.$$

Hence,

$$\sup_{\varphi\in\mathcal{G}} \left| \int_0^{\frac{1}{n}} \varphi_j' \circ F_j^{-1} \frac{B_{n,j}}{f_j \circ F_j^{-1}} \left\{\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)\right\} \right| \to 0$$

in probability and similarly for the right tail. Now, for every $t \in (0,1)$ we have

$$\varphi_j \circ F_{n,j}^{-1}(t) - \varphi_j \circ F_j^{-1}(t) = \varphi_j'\{K_{n,\varphi_j}(t)\}\{F_{n,j}^{-1}(t) - F_j^{-1}(t)\} \tag{A.4}$$

for some $K_{n,\varphi_j}(t)$ between $F_{n,j}^{-1}(t)$ and $F^{-1}(t)$. Therefore, recall (A.4), to prove (A.2) it suffices to show that

$$\sup_{\varphi\in\mathcal{G}} \left| \int_{1/n}^{1-1/n} \varphi_j'\{F_j^{-1}(t)\} \frac{B_{n,j}(t)}{f_j\{F_j^{-1}(t)\}} \left[\varphi_j\{F_j^{-1}(t)\} - F_B^{-1}(\varphi)(t)\right] dt \right.$$
$$\left. - \int_{1/n}^{1-1/2} \varphi_j'\{K_{n,\varphi_j}(t)\} \frac{\rho_{n,j}(t)}{f_j\{F_j^{-1}(t)\}} \left[\varphi_j\{F_j^{-1}(t)\} - F_B^{-1}(\varphi)(t)\right] dt \right| \to 0 \tag{A.5}$$

in probability. To check it, we take $\nu \in (0, 1/2)$ in (A.1) to get

$$\int_{1/n}^{1-1/n} \frac{|\rho_{n,j}(t) - B_{n,j}(t)|}{f_j\{F_j^{-1}(t)\}} \sup_{\varphi\in\mathcal{G}} \left|\varphi_j\{F_j^{-1}(t)\} - F_B^{-1}(\varphi)(t)\right| dt$$

$$\leq n^{\nu-1/2} O_P(1) \int_{1/n}^{1-1/n} \frac{\{t(1-t)\}^\nu}{f_k\{F_k^{-1}(t)\}} \sup_{\varphi\in\mathcal{G}} \left|\varphi_j\{F_j^{-1}(t)\} - F_B^{-1}(\varphi)(t)\right| dt \to 0 \quad \text{(A.6)}$$

in probability using dominated convergence and (iii) of Lemma 8. We observe next that, for each $t \in (0, 1)$,

$$\sup_{\varphi_j\in\mathcal{G}_j} |K_{n,\varphi_j}(t) - F_j^{-1}(t)| \to 0$$

a.s., since $K_{n,\varphi_j}(t)$ lies between $F_{n,j}^{-1}(t)$ and $F_j^{-1}(t)$. Therefore, using (12) we see that

$$\sup_{\varphi_j\in\mathcal{G}_j} |\varphi_j'\{K_{n,\varphi_j}(t)\} - \varphi_j'\{F_j^{-1}(t)\}| \to 0$$

a.s. while, on the other hand, $\sup_{\varphi_j\in\mathcal{G}_j} |\varphi_j'\{K_{n,\varphi_j}(t)\} - \varphi_j'\{F_j^{-1}(t)\}| \leq 2K_j$. But then, by dominated convergence we get that

$$\mathrm{E}\Big[\sup_{\varphi_j\in\mathcal{G}_j} \left|\varphi_j'\{K_{n,\varphi_j}(t)\} - \varphi_j'\{F_j^{-1}(t)\}\right|^2 \Big] \to 0.$$

Since by (iii) of Lemma 8 we have that $t \mapsto \sqrt{t(1-t)}/[f_j\{F_j^{-1}(t)\}] \sup_{\varphi\in\mathcal{G}} |\varphi_j\{F_j^{-1}(t)\} - F_B^{-1}(\varphi)(t)|$ is integrable we conclude that

$$\mathrm{E} \sup_{\varphi\in\mathcal{G}} \int_{1/n}^{1-1/n} \left|\varphi_j'\{K_{n,\varphi_j}(t)\} - \varphi_j'\{F_j^{-1}(t)\}\right| \frac{|B_{n,j}(t)|}{f_j\{F_j^{-1}(t)\}} \left|\varphi_j\{F_j^{-1}(t)\} - F_B^{-1}(\varphi)(t)\right| dt$$

tends to 0 as $n \to \infty$ and, consequently,

$$\sup_{\varphi\in\mathcal{G}} \int_{1/n}^{1-1/n} \left|\varphi_j'\{K_{n,\varphi_j}(t)\} - \varphi_j'\{F_j^{-1}(t)\}\right| \frac{|B_{n,j}(t)|}{f_j\{F_j^{-1}(t)\}} \left|\varphi_j\{F_j^{-1}(t)\} - F_B^{-1}(\varphi)(t)\right| dt$$

vanishes in probability. Combining this fact with (A.6) we prove (A.5) and, as a consequence, (A.2). Finally, observe that for every integer $n \geq 1$, C has the same law as $\hat{C}_n$. This completes the proof. □

Proof of Theorem 4. From Skohorod Theorem; see, e.g., Theorem 1.10.4 in [33], we know that there exists on some probability space versions of C_n and C for which convergence of C_n to C holds almost surely. From now on, we place us on this space and observe that

$$\sqrt{n}\,\{A_n(\mathcal{G}) - A(\mathcal{G})\} \leq \sqrt{n} \inf_\Gamma U_n - \sqrt{n} \inf_\Gamma U = \inf_{\varphi\in\Gamma} C_n(\varphi). \quad \text{(A.7)}$$

Furthermore, if we consider the (a.s.) compact set $\Gamma_n = \{\varphi \in \mathcal{G} : U(\varphi) \leq \inf_\mathcal{G} U + 2\|C_n\|_\infty/\sqrt{n}\}$, then, if $\varphi \notin \Gamma_n$, $U_n(\varphi) \geq \inf_\mathcal{G} U + \|C_n\|_\infty/\sqrt{n}$, while if $\varphi \in \Gamma$, then, $U_n(\varphi) \leq \inf_\mathcal{G} U + \|C_n\|_\infty/\sqrt{n}$. Thus, necessarily, $\inf_\mathcal{G} U_n = \inf_{\Gamma_n} U_n = \inf_{\Gamma_n}(U_n - U + U) \geq \inf_{\Gamma_n}(U_n - U) + \inf_{\Gamma_n} U = \inf_{\Gamma_n}(U_n - U) + \inf_\Gamma U$. Together with (A.7), this entails

$$\inf_{\varphi\in\Gamma_n} C_n(\varphi) \leq \sqrt{n}\,\{A_n(\mathcal{G}) - A(\mathcal{G})\} \leq \inf_{\varphi\in\Gamma} C_n(\varphi)$$

Note that for the versions that we are considering $\|C_n - C\|_\infty \to 0$ almost surely. In particular, this implies that $\inf_\Gamma C_n \to \inf_\Gamma C$ almost surely. Hence, the proof will be complete if we show that almost surely,

$$\inf_{\Gamma_n} C_n \to \inf_\Gamma C. \quad \text{(A.8)}$$

To check this last point, consider a sequence $\varphi_n \in \Gamma_n$ such that $C_n(\varphi_n) \leq \inf_{\Gamma_n} C_n + 1/n$. By compactness of $\mathcal{G}$, taking subsequences if necessary, $\varphi_n \to \varphi_0$ for some $\varphi_0 \in \mathcal{G}$. Continuity of U yields $U(\varphi_n) \to U(\varphi_0)$ and as a consequence, that $U(\varphi_0) \leq \inf_\mathcal{G} U$, i.e., $\varphi_0 \in \Gamma$ almost surely, Furthermore,

$$\left|C_n(\varphi_n) - C(\varphi_0)\right| \leq \|C_n - C\|_\infty + |C(\varphi_n) - C(\varphi_0)| \to 0.$$

This shows that

$$\liminf_{\Gamma_n} \inf C_n \geq C(\varphi_0) \geqslant \inf_{\Gamma} C$$

and yields (A.8). This completes the proof. □

Proof of Corollary 5. In Theorem 3, take $\mu'_j = \mu_{n,j}$. Then, writing $\mathcal{L}^*$ for the conditional law given the $X_{i,j}$, the result of Theorem 3 reads

$$W_2^2\Big[\mathcal{L}[\{A_{m_n}(\mathcal{G})\}^{1/2}], \mathcal{L}^*[\{A^*_{m_n}(\mathcal{G})\}^{1/2}]\Big] \leq L^2 \frac{1}{J}\sum_{j=1}^{J} W_2^2(\mu_j, \mu_{n,j}),$$

with $L = \sup_{\varphi\in\mathcal{G}} \|\varphi'_j\|_\infty < \infty$. Since $W_r\{\mathcal{L}(aX+b), \mathcal{L}(aY+b)\} = aW_r\{\mathcal{L}(X), \mathcal{L}(Y)\}$ for $a > 0$, $b \in \mathbb{R}$, the latter bound gives

$$W_2^2\Big[\mathcal{L}[\sqrt{m_n}[\{A_{m_n}(\mathcal{G})\}^{1/2} - \{A(\mathcal{G})\}^{1/2}]], \mathcal{L}^*[\sqrt{m_n}[\{A^*_{m_n}(\mathcal{G})\}^{1/2} - \{A(\mathcal{G})\}^{1/2}]]\Big] \leq L^2 \frac{m_n}{\sqrt{n}}\frac{1}{J}\sum_{j=1}^{J} \sqrt{n}\, W_2^2(\mu_j, \mu_{n,j}).$$

As noted in the proof of Theorem 4, the assumptions imply that $\sqrt{n}\, W_2^2(\mu_j, \mu_{n,j})$ vanishes in probability. Also, Theorem 4 and the delta method yield that

$$\sqrt{m_n}\Big[\{A_{m_n}(\mathcal{G})\}^{1/2} - \{A(\mathcal{G})\}^{1/2}\Big] \rightsquigarrow \frac{1}{2\{A(\mathcal{G})\}^{1/2}}\gamma,$$

with γ the limiting law there, which, combined to the above bound, shows that

$$\sqrt{m_n}\Big[\{A^*_{m_n}(\mathcal{G})\}^{1/2} - \{A(\mathcal{G})\}^{1/2}\Big] \rightsquigarrow \frac{1}{2\{A(\mathcal{G})\}^{1/2}}\gamma$$

in probability. A further use of the delta method yields

$$\sqrt{m_n}\Big\{A^*_{m_n}(\mathcal{G}) - A(\mathcal{G})\Big\} \rightsquigarrow \gamma$$

in probability. The result follows now from Lemma 1 in [23]. □

Proof of Theorem 6. We assume for simplicity that $p = 1$. The general case follows with straightforward changes. Let us observe that

$$U_n(\theta) = \frac{1}{J}\sum_{j=1}\int_0^1 \Big\{\psi_j(\theta_j, G_{n,j}^{-1}) - \frac{1}{J}\sum_{k=1}^{J}\psi_k(\theta_k, G_{n,k}^{-1})\Big\}^2,$$

with $G_{n,j}$ the empirical distribution function on the $\varepsilon_{i,j}$s, which are iid G. A similar expression, replacing $G_{n,j}$ with G is valid for $U(\theta)$. Then (17) implies that $\sup_\theta |U_n(\theta) - U(\theta)| \to 0$, from which, recall (14), it follows that $\hat{\theta}_n \to \theta^*$ in probability. Note that the second part in Assumption (17) is a technical condition which ensures that, when considering a Taylor expansion in the integral of $U_n(\theta)$, the remainder term in $\psi'_j(\lambda, H_{n,j}^{-1}) - \psi'_j(\lambda, G_j^{-1})$ for any $H_{n,j}^{-1}$ lying between $G_{n,j}^{-1}$ and G_j^{-1} (obtained through a Taylor expansion) goes uniformly to zero.

From (15) we have that U_n is a C^2 function whose derivatives can be computed by differentiation under the integral sign. This implies that

$$D_j U_n(\theta) = \frac{2}{J}\int_0^1 D\psi_j(\theta_j, G_{n,j}^{-1})\Big\{\psi_j(\theta_j, G_{n,j}^{-1}) - \frac{1}{J}\sum_{k=1}^{J}\psi_k(\theta_k, G_{n,k}^{-1})\Big\},$$

$$D_{p,q}U_n(\theta) = -\frac{2}{J^2}\int_0^1 D\psi_p(\theta_p, G_{n,p}^{-1})D\psi_q(\theta_q, G_{n,q}^{-1}),\ p \neq q \tag{A.9}$$

and

$$D_{p,p}U_n(\theta) = \frac{2}{J}\int_0^1 D^2\psi_p(\theta_p, G_{n,p}^{-1})\Big\{\psi_j(\theta_j, G_{n,j}^{-1}) - \frac{1}{J}\sum_{k=1}^{J}\psi_k(\theta_k, G_{n,k}^{-1})\Big\} + \frac{2(J-1)}{J^2}\int_0^1 \left\{D\psi_p(\theta_p, G_{n,p}^{-1})\right\}^2.$$

Using also (16) we obtain similar expressions for the derivatives of $U(\theta)$, replacing everywhere $G_{n,j}^{-1}$ with G^{-1}. We write $DU_n(\theta) = \{D_jU_n(\theta)\}_{1\le j\le J}$, $DU(\theta) = \{D_jU(\theta)\}_{1\le j\le J}$ for the gradients and $\Sigma_n(\theta) = \{D_{p,q}U_n(\theta)\}_{1\le p,q\le J}$, $\Sigma(\theta) = \{D_{p,q}U(\theta)\}_{1\le p,q\le J}$for the Hessians of U_n and U. Note that $\Sigma^* = \Sigma(\theta^*)$ is assumed to be invertible.

We write now $\rho_{n,j}$ for the quantile process based on the $\varepsilon_{i,j}$s. Observe that (18) ensures that we can assume, without loss of generality, that there exist independent Brownian bridges, $B_{n,j}$, satisfying (A.1). Now, recalling that $\psi_j(\theta_j^*, x) = x$, we see that

$$\sqrt{n}\, D_jU_n(\theta^*) = \frac{2}{J}\int_0^1 D\psi_j\{\theta_j^*, G_{n,j}^{-1}(t)\}\,\frac{\rho_{n,j}(t) - \sum_{k=1}^{J}\rho_{n,k}(t)/J}{g\{G^{-1}(t)\}}dt.$$

Now, using (16) and arguing as in the proof of Theorem 4, we conclude that

$$\left|\int_0^1 D\psi_j\{\theta_j^*, G_{n,j}^{-1}(t)\}\frac{\rho_{n,k}(t)}{g\{G^{-1}(t)\}}dt - \int_0^1 D\psi_j\{\theta_j^*, G^{-1}(t)\}\frac{B_{n,k}(t)}{g\{G^{-1}(t)\}}dt\right| \to 0$$

in probability and, consequently,

$$\left|\sqrt{n}\, D_jU_n(\theta^*) - \frac{2}{J}\int_0^1 D\psi_j\{\theta_j^*, G^{-1}(t)\}\frac{B_{n,j}(t) - \sum_{k=1}^{J}B_{n,k}(t)/J}{g\{G^{-1}(t)\}}dt\right| \to 0 \tag{A.10}$$

in probability.

A further Taylor expansion of D_jU_n around θ^* shows that for some $\tilde\theta_j^n$ between $\hat\theta_n$ and θ^* we have

$$D_jU_n(\hat\theta_n) = D_jU_n(\theta^*) + (D_{1j}U_n(\tilde\theta_j^n), \ldots, D_{Jj}^2U_n(\tilde\theta_j^n)) \times (\hat\theta_n - \theta^*)$$

and because $\hat\theta_n$ is a zero of DU_n, we obtain

$$-D_jU_n(\theta^*) = (D_{1j}U_n(\tilde\theta_j^n), \ldots, D_{Jj}U_n(\tilde\theta_j^n)) \times (\hat\theta_n - \theta^*).$$

Writing $\tilde\Sigma_n$ for the $J\times J$ matrix whose Jth row equals $(D_{1j}U_n(\tilde\theta_j^n), \ldots, D_{Jj}U_n(\tilde\theta_j^n))$, with $j \in \{1, \ldots, J\}$, we can rewrite the last expansion as

$$-\sqrt{n}\, DU_n(\theta^*) = \tilde\Sigma_n\sqrt{n}(\hat\theta_n - \theta^*). \tag{A.11}$$

Now, recalling (A.9), assumptions (15) and (16) yield that $\tilde\Sigma_n \to \Sigma^* = \Sigma(\theta^*)$ in probability. As a consequence, (A.11) and (A.10) together with Slutsky's Theorem complete the proof of the second claim.

Finally, for the proof of the last claim, since $DU_n(\hat\theta_n) = 0$, a Taylor expansion around $\hat\theta_n$ shows that

$$nU_n(\theta^*) - nU_n(\hat\theta_n) = \frac{1}{2}\{\sqrt{n}\,(\hat\theta_n - \theta^*)\}^\top\Sigma(\tilde\theta_n)\{\sqrt{n}\,(\hat\theta_n - \theta^*)\}$$

for some $\tilde\theta_n$ between $\hat\theta_n$ and θ^*. Arguing as above we see that $\Sigma(\tilde\theta_n) \to \Sigma^*$ in probability. Hence, to complete the proof if suffices to show that

$$nU_n(\theta^*) - \frac{1}{J}\sum_{j=1}^{k}\int_0^1 \frac{\{B_{n,j}(t) - \sum_{k=1}^{J}B_{n,k}(t)/J\}^2}{g\{G^{-1}(t)\}^2}dt \to 0$$

in probability. Since

$$nU_n(\theta^*) = \frac{1}{J}\sum_{j=1}^{k}\int_0^1 \frac{\{\rho_{n,j}(t) - \sum_{k=1}^{J}\rho_{n,k}(t)/J\}^2}{g\{G^{-1}(t)\}^2}dt,$$

this amounts to proving that

$$\int_0^1 \{\rho_{n,j}(t) - B_{n,j}(t)\}^2 / g\{G^{-1}(t)\}^2 dt \to 0$$

in probability. Taking $\nu \in (0, 1/2)$ in (A.1), we see that

$$\int_{1/n}^{1-1/n} \{\rho_{n,j}(t) - B_{n,j}(t)\}^2 \, g\{G^{-1}(t)\}^2 dt \le O_P(1) \frac{1}{n^{1-2\nu}} \int_{1/n}^{1-1/n} \{t(1-t)\}^{2\nu} / g\{G^{-1}(t)\}^2 \to 0,$$

using condition (19) and dominated convergence. From (19) we also see that

$$\int_{1-1/n}^1 B_{n,j}(t)^2 / g\{G^{-1}(t)\}^2 dt \to 0$$

in probability. Condition (19) implies also that

$$\int_{1-1/n}^1 \rho_{n,j}(t)^2 / g\{G^{-1}(t)\}^2 dt \to 0$$

in probability; see [30]. Similar considerations apply to the left tail and complete the proof. □

Proof of Corollary 7. Writing $\mathcal{L}^*$ for the conditional law given the $X_{i,j}$s, we see from Theorem 3 that

$$W_2^2\Big[\mathcal{L}[\sqrt{m_n}\{A_{m_n}(\Theta)\}^{1/2}], \mathcal{L}^*[\sqrt{m_n}\{A^*_{m_n}(\Theta)\}^{1/2}]\Big] \le L \frac{m_n}{n} \frac{1}{J} \sum_{j=1}^J n W_2^2(\mu, \tilde{\mu}_{n,j}),$$

where $L = \sup_{\lambda,x,j} \psi_j'(\lambda, x)$, μ denotes the law of the errors, $\varepsilon_{i,j}$, and $\tilde{\mu}_{n,j}$ the empirical distribution function on $\varepsilon_{1,j}, \ldots, \varepsilon_{n,j}$. Note that $L < \infty$ by (17), while $nW_2^2(\mu, \tilde{\mu}_{n,j}) = O_P(1)$ as in the proof of Theorem 6. Hence, we conclude that

$$m_n A^*_{m_n}(\Theta) \rightsquigarrow \frac{1}{J} \sum_{j=1}^J \int_0^1 \left(\tilde{B}_j / g \circ G^{-1}\right)^2 - Y^\top \Sigma^{-1} Y/2$$

in probability. The conclusion now follows from Lemma 1 in [23]. □

If centering were necessary and we had (21) rather than the limit in Theorem 6, we could adapt the last argument as follows. If A and B are positive random variables, then $\mathrm{E}|A - B| \le \mathrm{E}(A^{1/2} - B^{1/2})^2 + 2\{\mathrm{E}A\,\mathrm{E}(A^{1/2} - B^{1/2})^2\}^{1/2}$. We can apply this bound to (an optimal coupling of) $m_n A_{m_n}(\Theta)$ and $m_n A^*_{m_n}(\Theta)$. Now if the errors have a log-concave distribution then $n\mathrm{E}W_2^2(\mu, \tilde{\mu}_{n,j}) = O(\ln n)$; see Corollary 6.12 in [7]. We conclude that

$$W_1[\mathcal{L}\{m_n A_{m_n}(\Theta) - c_{m_n}\}, \mathcal{L}^*\{m_n A^*_{m_n}(\Theta) - c_{m_n}\}] = W_1[\mathcal{L}\{m_n A_{m_n}(\Theta)\}, \mathcal{L}^*\{m_n A^*_{m_n}(\Theta)\}]$$

vanishes in probability if $m_n = O(n^\rho)$ for some $\rho \in (0, 1)$. As a consequence,

$$m_n A^*_{m_n}(\Theta) - c_{m_n} \rightsquigarrow \frac{1}{J} \sum_{j=1}^J \int_0^1 \frac{\tilde{B}_j^2 - E\tilde{B}_j^2}{(g \circ G^{-1})^2} - \frac{1}{2} Y^\top \Sigma^{-1} Y$$

in probability. □

Ackowledgements

The authors would like to thank the Editor for his careful reading of the manuscript and his suggestions that improved the paper.

E. del Barrio's research has been partially supported by Spanish Ministerio de Economía y Competitividad, grant MTM2017-86061-C2-1-P, and by Consejería de Educación de la Junta de Castilla y León and FEDER, grants VA005P17 and VA002G18

Table 1: Simulations under $\mathcal{H}_0$.

J	n	$m_n = n^{0.6}$	$m_n = n^{0.7}$	$m_n = n^{0.8}$	$m_n = n^{0.9}$	$m_n = n^{0.95}$	$m_n = n$
	50	0.144	0.079	0.038	0.046	0.041	0.03
	100	0.148	0.067	0.07	0.05	0.04	0.033
	200	0.129	0.085	0.068	0.043	0.037	0.044
2	500	0.138	0.089	0.05	0.048	0.035	0.036
	1000	0.127	0.086	0.063	0.055	0.039	0.032
	2000	0.129	0.104	0.071	0.048	0.043	0.038
	5000	0.039	0.042	0.041	0.049	0.043	0.055
	50	0.295	0.194	0.115	0.078	0.054	0.034
	100	0.273	0.163	0.089	0.053	0.034	0.039
	200	0.238	0.15	0.077	0.054	0.047	0.031
3	500	0.226	0.122	0.07	0.057	0.042	0.029
	1000	0.217	0.107	0.092	0.069	0.042	0.035
	2000	0.221	0.128	0.077	0.053	0.043	0.035
	5000	0.205	0.145	0.082	0.06	0.025	0.047
	50	0.659	0.428	0.281	0.129	0.111	0.081
	100	0.583	0.337	0.192	0.104	0.083	0.053
	200	0.538	0.281	0.159	0.081	0.078	0.029
5	500	0.449	0.267	0.138	0.063	0.056	0.04
	1000	0.415	0.238	0.129	0.064	0.051	0.037
	2000	0.354	0.212	0.115	0.06	0.053	0.032
	5000	0.322	0.203	0.108	0.057	0.061	0.039
	50	0.996	0.971	0.873	0.702	0.553	0.456
	100	0.994	0.902	0.708	0.433	0.33	0.226
	200	0.958	0.802	0.521	0.247	0.184	0.119
10	500	0.914	0.663	0.388	0.149	0.093	0.063
	1000	0.864	0.532	0.286	0.119	0.084	0.046
	2000	0.813	0.473	0.239	0.103	0.063	0.051
	5000	0.756	0.449	0.217	0.088	0.061	0.041

Table 2: Power of the test for $\gamma =_d \mathcal{E}(1)$.

J	n	$m_n = n^{0.6}$	$m_n = n^{0.7}$	$m_n = n^{0.8}$	$m_n = n^{0.9}$	$m_n = n^{0.95}$	$m_n = n$
2	50	0.961	0.919	0.897	0.864	0.829	0. 838
	100	1	0.998	0.998	0.995	0.994	0.993
	200	1	1	1	1	1	1
	500	1	1	1	1	1	1
	1000	1	1	1	1	1	1
	2000	1	1	1	1	1	1
	5000	1	1	1	1	1	1
3	50	0.987	0.971	0.97	0.953	0.939	0.91
	100	1	1	0.999	1	0.999	0.999
	200	1	1	1	1	1	1
	500	1	1	1	1	1	1
	1000	1	1	1	1	1	1
	2000	1	1	1	1	1	1
	5000	1	1	1	1	1	1
5	50	1	0.996	0.988	0.976	0.971	0.955
	100	1	1	1	1	1	1
	200	1	1	1	1	1	1
	500	1	1	1	1	1	1
	1000	1	1	1	1	1	1
	2000	1	1	1	1	1	1
	5000	1	1	1	1	1	1
10	50	1	1	1	1	0.996	0.985
	100	1	1	1	1	1	1
	200	1	1	1	1	1	1
	500	1	1	1	1	1	1
	1000	1	1	1	1	1	1
	2000	1	1	1	1	1	1
	5000	1	1	1	1	1	1

Table 3: Power of the test $\gamma \overset{d}{=} \text{Laplace}(0,1)$.

J	n	$m_n = n^{0.6}$	$m_n = n^{0.7}$	$m_n = n^{0.8}$	$m_n = n^{0.9}$	$m_n = n^{0.95}$	$m_n = n$
2	50	0.426	0.33	0.3	0.241	0.223	0.163
	100	0.658	0.534	0.468	0.365	0.361	0.3
	200	0.855	0.824	0.751	0.665	0.613	0.602
	500	0.998	0.998	0.993	0.982	0.965	0.962
	1000	1	1	1	1	0.999	1
	2000	1	1	1	1	1	1
	5000	1	1	1	1	1	1
3	50	0.657	0.533	0.422	0.331	0.282	0.223
	100	0.831	0.708	0.586	0.514	0.461	0.377
	200	0.946	0.915	0.841	0.778	0.709	0.661
	500	1	0.998	0.997	0.994	0.989	0.977
	1000	1	1	1	1	1	1
	2000	1	1	1	1	1	1
	5000	1	1	1	1	1	1
5	50	0.895	0.741	0.633	0.471	0.394	0.333
	100	0.936	0.874	0.728	0.623	0.519	0.443
	200	0.994	0.947	0.903	0.847	0.786	0.696
	500	1	1	1	0.996	0.992	0.985
	1000	1	1	1	1	1	1
	2000	1	1	1	1	1	1
	5000	1	1	1	1	1	1
10	50	1	0.997	0.97	0.875	0.79	0.703
	100	0.997	0.985	0.949	0.854	0.765	0.643
	200	1	0.996	0.968	0.924	0.859	0.789
	500	1	1	1	0.996	0.996	0.975
	1000	1	1	1	1	1	0.999
	2000	1	1	1	1	1	1
	5000	1	1	1	1	1	1

Table 4: Power of the test $\gamma \overset{d}{=} t_{(3)}$.

I	n	$m_n = n^{0.6}$	$m_n = n^{0.7}$	$m_n = n^{0.8}$	$m_n = n^{0.9}$	$m_n = n^{0.95}$	$m_n = n$
2	50	0.566	0.445	0.429	0.352	0.321	0.307
	100	0.775	0.704	0.647	0.576	0.503	0.454
	200	0.942	0.927	0.882	0.833	0.771	0.697
	500	1	0.997	0.995	0.991	0.989	0.957
	1000	1	1	1	1	1	0.986
	2000	1	1	1	1	1	0.999
	5000	1	1	1	1	1	0.997
3	50	0.745	0.653	0.546	0.46	0.402	0.349
	100	0.881	0.821	0.738	0.65	0.592	0.563
	200	0.98	0.958	0.928	0.891	0.873	0.794
	500	1	1	0.999	0.997	0.997	0.978
	1000	1	1	1	1	1	0.995
	2000	1	1	1	1	1	1
	5000	1	1	1	1	1	1
5	50	0.91	0.813	0.682	0.593	0.525	0.45
	100	0.972	0.909	0.822	0.751	0.686	0.621
	200	0.995	0.984	0.967	0.915	0.887	0.836
	500	1	1	1	0.999	0.999	0.995
	1000	1	1	1	1	1	1
	2000	1	1	1	1	1	1
	5000	1	1	1	1	1	1
10	50	1	0.997	0.953	0.894	0.827	0.758
	100	0.999	0.993	0.969	0.907	0.862	0.79
	200	1	0.998	0.995	0.961	0.941	0.903
	500	1	1	1	1	0.998	0.988
	1000	1	1	1	1	1	0.998
	2000	1	1	1	1	1	0.999
	5000	1	1	1	1	1	1

Table 5: Power of the test $\gamma \overset{d}{=} t_{(4)}$.

I	n	$m_n = n^{0.6}$	$m_n = n^{0.7}$	$m_n = n^{0.8}$	$m_n = n^{0.9}$	$m_n = n^{0.95}$	$m_n = n$
2	50	0.398	0.353	0.292	0.207	0.182	0.183
	100	0.623	0.52	0.429	0.341	0.29	0.228
	200	0.826	0.717	0.65	0.589	0.526	0.41
	500	0.989	0.978	0.954	0.928	0.878	0.787
	1000	1	1	0.999	1	0.984	0.955
	2000	1	1	1	1	1	0.985
	5000	1	1	1	1	1	0.993
3	50	0.634	0.495	0.4	0.295	0.263	0.222
	100	0.756	0.666	0.56	0.465	0.399	0.336
	200	0.914	0.859	0.778	0.663	0.602	0.521
	500	0.998	0.989	0.985	0.972	0.928	0.868
	1000	1	1	1	1	0.999	0.963
	2000	1	1	1	1	1	0.989
	5000	1	1	1	1	1	1
5	50	0.851	0.709	0.583	0.426	0.359	0.316
	100	0.919	0.825	0.668	0.546	0.493	0.316
	200	0.959	0.908	0.842	0.738	0.684	0.578
	500	1	0.997	0.994	0.973	0.934	0.888
	1000	1	1	1	1	0.999	0.968
	2000	1	1	1	1	1	1
	5000	1	1	1	1	1	0.999
10	50	1	0.986	0.941	0.813	0.774	0.653
	100	1	0.988	0.925	0.806	0.738	0.606
	200	1	0.991	0.948	0.854	0.813	0.679
	500	1	1	0.998	0.985	0.954	0.886
	1000	1	1	1	1	0.997	0.949
	2000	1	1	1	1	1	0.974
	5000	1	1	1	1	1	0.995

References

[1] M. Agueh, G. Carlier, Barycenters in the Wasserstein space, SIAM J. Math. Anal. 43 (2011) 904–924.

[2] M. Agulló-Antolín, J.A. Cuesta-Albertos, H. Lescornel, J.-M. Loubes, A parametric registration model for warped distributions with Wasserstein's distance, J. Multivariate Anal. 135 (2015) 117–130.

[3] S. Allassonnière, Y. Amit, A. rouvé, Towards a coherent statistical framework for dense deformable template estimation, J. R. Stat. Soc. Ser. B (Stat. Methodol.) 69 (2007) 3–29.

[4] P.C. Álvarez-Esteban, E. del Barrio, J.A. Cuesta-Albertos, C. Matrán, Trimmed comparison of distributions, J. Amer. Statist. Assoc. 103 (2008) 697–704.

[5] Y. Amit, U. Grenander, M. Piccioni, Structural image restoration through deformable template, J. Amer. Statist. Assoc. 86 (1991) 376–387.

[6] J. Bigot, T. Klein, Characterization of barycenters in the Wasserstein space by averaging optimal transport maps, ESAIM: Probability and Statistics, in press (2018).

[7] S. Bobkov, M. Ledoux, One-dimensional empirical measures, order statistics and Kantorovich transport distances, Memoirs of the American Mathematical Society, in press (2018).

[8] E. Boissard, T. Le Gouic, J.-M. Loubes, Distribution's template estimate with Wasserstein metrics, Bernoulli 21 (2015) 740–759.

[9] B.M. Bolstad, R.A. Irizarry, M. Åstrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, Bioinformatics 19 (2003) 185–193.

[10] V. Chernozhukov, A. Galichon, M. Hallin, M. Henry, Monge–Kantorovich depth, quantiles, ranks, and signs, Ann. Statist. 45 (2017) 223–256.

[11] O. Collier, A.S. Dalalyan, Curve registration by nonparametric goodness-of-fit testing, J. Statist. Plann. Inf. 162 (2015) 20–42.

[12] M. Csörgő, Quantile processes with statistical applications, CBMS-NSF Regional Conference Series in Applied Mathematics 42, SIAM, 1983.

[13] M. Csörgő, L. Horváth, Weighted Approximations in Probability and Statistics, Wiley, New York, 1993.

[14] M. Cuturi, A. Doucet, Fast computation of Wasserstein barycenters, Proceedings of the International Conference on Machine Learning 2014, JMLR W&CP 32 (2014) 685–693.

[15] C. Czado, A. Munk, Assessing the similarity of distributions–finite sample performance of the empirical Mallows distance, J. Statist. Comput. Simul. 60 (1998) 319–346.

[16] E. del Barrio, P. Deheuvels, S. van de Geer, Lectures on Empirical Processes: Theory and Statistical Applications, European Mathematical Society, Zürich, Switzerland, 2007.

[17] E. del Barrio, E. Giné, F. Utzet, Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances, Bernoulli 11 (2005) 131–189.

[18] E. Del Barrio, J.-M. Loubes, Central Limit Theorem for empirical transportation cost in general dimension. ArXiv e-prints 1705.01299, 2017.

[19] J. Ebert, V. Spokoiny, A. Suvorikova, Construction of Non-asymptotic Confidence Sets in 2-Wasserstein Space ArXiv preprint 1703.03658, 2017.

[20] G. Freitag, A. Munk, On Hadamard differentiability in k-sample semiparametric models, with applications to the assessment of structural relationships, J. Multivariate Anal. 94 (2005) 123–158.

[21] S. Gallón, J.-M. Loubes, E. Maza, Statistical properties of the quantile normalization method for density curve alignment, Math. Biosci. 242 (2013) 129–142.

[22] F. Gamboa, J.-M. Loubes, E. Maza, Semi-parametric estimation of shifts, Electron. J. Statist. 1 (2007) 616–640.

[23] A. Janssen, T. Pauls, How do bootstrap and permutation tests work? Ann. Statist. 31 (2003) 768–806.

[24] T. Le Gouic, J.-M. Loubes, Existence and consistency of Wasserstein barycenters, Prob. Theory Related Fields 168 (2017) 901–917.

[25] A. Munk, C. Czado, Nonparametric validation of similar distributions and assessment of goodness of fit, J. R. Stat. Soc. Ser. B (Stat. Methodol.) 60 (1998) 223–241.

[26] S.T. Rachev, The Monge–Kantorovich problem on mass transfer and its applications in stochastics, Teor. Veroyatnost. i Primenen. 29 (1984) 625–653.

[27] B.S. Rajput, Gaussian measures on L_p spaces, $1 \leq p < \infty$, J. Multivariate Anal. 2 (1972) 382–403.

[28] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, 2nd Ed., Springer, New York, 2005.

[29] T. Rippl, A. Munk, A. Sturm, Limit laws of the empirical Wasserstein distance: Gaussian distributions, arXiv preprint arXiv:1507.04090, 2015.

[30] R. Samworth, O. Johnson, Convergence of the empirical process in Mallows distance, with an application to bootstrap performance. arXiv:math/0406603, 2004.

[31] M. Sommerfeld, A. Munk, Inference for empirical Wasserstein distances on finite spaces, J. R. Stat. Soc. B (Stat. Methodol.) 80 (2018) 219–238.

[32] A.W. van der Vaart, Asymptotic Statistics, Cambridge University Press, 2000.

[33] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, Springer, New York, 1996.

[34] C. Villani, Optimal Transport: Old and New, Springer Verlag, New York, 2009.